Naming identifiers Allowed: letters, digits, period () and underscore (). Must start with letter or period. If starts with period, cannot be followed by digit. Cannot use keywords such as TRUE, IF, etc.			Operations & Data Structures AND: & OR: NOT: ! Missing Values: NA			sum() FUNCTION sum() is another aggregate function that can be used within summarise() or mutate(). Returns the total values of the vector passed as an argument. Ex. Calculate the sum of total assets of all firms by country of							
variable <- "value" typeof(variable) -> "logical"			Vectors			Ex . Calculate the sum of total assets of all firms by country of headquarters (loc) and fiscal year (fyear). Call this new variable tot at.							
typeontvariable) -> "togical"			Must be of same data type.			Retain all the observations in the original data.					new variable cot_at.		
Userul Facts			city <- c("Vanco	ouver", "Victoria'	", "Delta", "Surrey")	Canuseth	ne sum() fu	unction with n	nutate()	Your	result should look lik	e the following:	
As you create (i.e., assign) objects in an R session, the objects will appear in the Environment . Use ? to look for help in R. To install a package, do install.packages("name of the package")			<pre>length(x) > city <- c("Vancouver", "Victoria", "Delta", "Surrey") > length(city)</pre>			can use the sum() function with mutate().				i works (iii) Add Care (iii) (iii) (iii) i works Add Care Add Care works works works i works Mark Care Add Care works works works i works works works works works works works i works works works works works works works works i works Mark works			
Dataframes			[1] 4			filter(!is group by(f	.na(fyear year, loo	r), !is.na(c) %>%	loc)) %>%		8 003004 2003 AAR COMP 9 003004 2002 AAR COMP 38 003004 2003 AAR COMP	USA 718.199 29613647 USA 686.621 29960360 USA 799.292 32862366	
Datairairies			Subcotting	> expensive	e_city <- city[1:2]	mutate (tot	al_at = s	sum(at, na.	rm = TRUE))	8>8	11 003004 2304 AAR CORP 12 003004 2395 AAR CORP Showing 1 to 13 of 288,313 entries, 6 total columns	USA 732.239 36200138 USA 978.819 36282897	
a table with columns representing variables and rows rep	presenting observation	ıs.	R IS ONE INDEX	XED. [1] "Vancou	aver" "Victoria"	select (gvk	ey, iyea:	r, conm, 10	oc, at, tota	group	rms located in the sar ped together as a sub	me loc and the sam oset, and then R retu	e fyear are irns the total
(think of a 2x2 matrix array, All vectors in data frame mus	t have the same numb	er of elems.)	Use - to remov	e elements, : to i	return range. can ass	ign subsets	to new veo	ctors.		ofat	for each subset – th	e country/year com	bination.
can also use data.frame(), a part of R's standard library			city[2] returns	"Victoria"						arrange() Fl		(a) an acified in .	
movies <- data_trame(city[1:3] return	s a vector "Vanc	ouver" "Victoria" "De	lta"				orders the r order.	ows by variable	(s) specified, in a	ascending
title ~ (C"Groundhog Day", "Mission: Impossible", "Inception", "Terminator 2", "Back to the Future"), year ~ c(193, 1996, 2010, 1991, 1985), duration ~ c(101, 110, 148, 137, 116), box.office ~ c(105, 438, 437, 521, 389),			city[c(1,2,4)] returns specified elements "vancouv city[-3] returns "Vancouver" "Victoria" "Surrey"			er", "Victoria", "Surrey" > expensive_city(2) <- "Toronto" > expensive_city				Ex. Arrange observations in companies in ascending order of fyear and descending			
rating = c(96, 66, 87, 93, 96)	es ×		Can update the	e value(s) of the s	subset of a vector.	[1] "Vancou	ver" "Toront	:0"		order of ch	within each fyea	ır.	
Environment History Connections Tutorial	title year duration 1	box_office ⁰ rating ⁰	Can also add ne	ew elements to a	an existing vector.	<pre>[1] 2 > expensive_ci</pre>	ty[3:4] <- c("	New York","Los Ang	eles")	ordered <	<- companies	%>%	1 0.00
Circle for the second s	Groundhog Day 1993 101 Mission: Impossible 1996 110	458 66	If you perform	an operation wit	th vectors the	<pre>> expensive_ci [1] "Vancouver"</pre>	ty "Toronto"	"New York"	"Los Angeles"	arrange	(fyear, desc(ch)) %>%	1 0/0
Bata 4	Inception 2010 148 Terminator 2 1991 137	837 87 521 93	operation will I	be applied to eac	ch element of the ve	ctor. > city_f	temp + 2			select(f	fyear, ch, gv	key, conm)	
movies 5 obs. of 5 variables	Back to the Future 1985 116	389 96	Function	What it returns		[1] 27 2	22 29 32			Unctied1* Orde Orde Orde Orde Orde	ered ×		(Q,)
get column values. name_or_data_frame\$name_or_colun	[1] 96 66 87 93 96	6	max(x)	the maximum va	lue of x					1 1994 18212 2 1994 17823	2764 005650 HITACHI LTD 1.301 015627 BANK TOKYO-MITSUI	ND4	
Suppose you are asked to create a vector whose element	s		min(x)	the minimum val	lue of x d maximum values of x					3 1994 17359 4 1994 16298	899 007114 PANASONIC CORP 8,743 019661 TOYOTA MOTOR CO	<i>w</i>	
are TRUE if each of the corresponding movies is longer th 2 hours, and FAI SE otherwise	[1] FALSE FALSE TRU	UE TRUE FALSE	length(x)	the number of el	ements in x					5 1994 15493 6 1994 13578	1.899 019349 SEMENS AG 1.000 002024 BANKAMERICA CORP	010	
create new column: name of dataframe\$new col <- mov	ies\$duration/60		sum(x)	the sum of the el	lements in x					7 1994 13377 8 1994 11442 9 1994 10619	7.898 013294 NATL WISTMINSTER 2.000 015208 FEDERAL HOME LOAN 8.000 012826 HANSON PLC	MORTG CORP	
movies\$hours <- movies\$duration/60	resputiation/00		mean(x)	the average value	e of the elements in x					10 1994 10083 11 1994 9669	1.699 015889 ANZ-AUSTRALIA & N 0.000 007471 MITSUI & CO LITD	DW ZEALD BK	
Consider = Consider = Public Town = Public Town = Public Public Towns = Public Towns = Public Town =			sd(x) var(x)	the standard dev	nation of x					Showing 1 to 11 of 232,362	t entries		
1 Grawdhog Day 1919 191 195 66 64.852 1.643131 2 Wassen mpossible 1959 118 4-58 66 1706 3.831333 3 Wassen mpossible 2309 118 4-58 66 1706 3.845647			sqrt(x)	the square root o	of x								
4 Jonnacure 2 1394 132 132 131 2.233331 8 Back to the Future 1398 116 189 96 54332 1.5931333	Data Wrangling	g	round(value_to_ro	ound, decimal_place	es_to_round_to) ex. roun	d(x/y, 3))		FOR	A FORMULA BEING IED:	IF THE RE	FERENCE IS:	IT CHANGES TO:	
NULL VALUES	The process of trans	sforming and m	apping data fror	m one raw data f	orm into another,			1	A B C	\$A\$1 (abs row)	solute column and absolut	e \$A\$1 (the reference absolute)	ce is
By default, if there is a missing > 1+NA > mean(x)	with the intent of m	aking the data r	more useful (e.g.	, for analytics).	,			2 3					
value, the mean will evaluate [1] NA [1] NA [1] NA	Packages used: dply	yr, magrittr ship	oped with tidyve	rse						A\$1 (relat	ive column and absolute r	ow) C\$1 (the reference	a is mixed)
to NA [1] NA [1] NA $>$ NA $ =$ TRUE	dplyr				MAGRITIR					SAI (abso A1 (relativ	olute column and relative r	 SA3 (the reference) C3 (the reference) 	is relative)
[1] NA	Function	now it works	od on their names		pipe operator								
na.rm() is a logical value indicating whether NA values should be ignored, default is FALSE	filter()	nicks observations	based on their value	•	functions (like select	ator (I.e., %> t() and filter	>%) is used r()) take the	i, the dplyr e data					
<pre>> mean(x, na.rm = TRUE)</pre>	mutate()	adds new variables	s that are functions o	f existing variables	from the previous s	tep as their	input.						
[1] 2.5	summarise()	reduces multiple va	alues to a summary s	statistics	companies %>% filt	er(loc=="CA	N") is iden	ntical to filter	(companies, l	oc=="CAN")			
is.na(x) indicates which elements are missing.	arrange()	changes the order	of rows		examples:								
> is.na(x)	group by()	allows a user to per	rform the above fund	ctions on a subset of			inner_jo	oin(datase	et_x, datas	et_y) FUN	NCTION		
[I] FALSE FALSE FALSE FALSE TRUE	NIVERSITY OF BRITISH COLUMBIA	the data		SAUDED SCHOOL OF BUSINESS									m x and y
<pre>sum(is.na(x)) adds up all the elements that</pre>							returns all	l rows from x	where there a	are matching	g values in y , and	all columns fro	min x anu y.
evaluate to 1 thus giving the number of missing	Loading rds files (sing	gular r objects)			9 (period)* - mock price, pear and -	-0	returns all	l rows from x	where there a	are matching	g values in y , and	all columns fro	Jill x allu y.
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames.	<pre>> companies <- readRD</pre>	gularrobjects) S("~/North Ameri	ican Stock Market	1994-2018.rds")	Destinal * mack_proc_war_and * Sec. 10 * mack_proc_war_and * model area_war_and *		returns all merged1 <	l rows from x <- inner_join(e	where there a example1,exar	are matching mple2)	g values in y , and	all columns fro	Jin x and y.
<pre>evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x))</pre>	Loading rds files (sing > companies <- readRD select(data_object_na	gular r objects) IS("~/North Ameri ame, variable_n	ican Stock Market	: 1994-2018.rds")	B Constant 1* Statut, price, pear, and - C C V Nor I And COSP UGA And I And COSP UGA And I And COSP UGA And	Q	returns all merged1 < To specify f	<pre>l rows from x <- inner_join(e this match more</pre>	where there a example1,exar e clearly (which i	are matching mple2) is good practic	g values in y, and	all columns fro	Jin x and y.
<pre>evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1</pre>	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variable	gular r objects) IS("~/North Ameri ame, variable_n les you specify	ican Stock Market name(s)) from the datase	: 1994-2018.rds") :t	9 testing prior, year, year, year 1 1 1 Filter 1 1 4 6 6 4 2 44 60 6 4 3 44 60 6 4 4 444 60 66 4 4 444 60 66 46 4 444 60 66 46 4 444 60 66 46	Apex prot_c Apex prot_c 1094 1.8.74999 1095 2.189798 1096 3.0.74985 1096 3.0.74985 1096 2.1.87500 1096 2.1.87500	returns all merged1 < To specify t you can do	I rows from x - inner_join(e) this match more the following:	where there a example1,exar e clearly (which i	mple2) is good practic	g values in y , and xe),	all columns fro	Jin x and y.
<pre>evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum((is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted</pre>	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variable price_yr_end <- select	gular r objects) IS("~/North Ameri ame, variable_n les you specify t(companies, co	ican Stock Market name(s)) from the datase onm, loc, tic, fyea	: 1994-2018.rds") :t ar, prcc_c)	9. Jonatol ** Instal procepany part = 1 11. ** * access 2 Acc COP* 3 Acc COP* 4 Acc COP* 4 Acc COP* 5 Acc COP* 6 Acc COP* 6 Acc COP* 6 Acc COP* 6 Acc COP* 7 Acc COP* 6 Acc COP* 7 Acc COP*	mm mm futur putc, c 1964 13.374699 1964 13.374699 1964 13.374699 1965 20.24985 1966 30.24985 1969 13.87500 1969 13.87500 1969 13.87500 2000 13.82500 2001 5.815000 2002 5.815000	returns all merged1 < To specify t you can do merged1 <- in	I rows from x - inner_join(et this match more the following: ner_join(example)	where there a example1,exar e clearly (which i ,example2, by = c(mple2) is good practic ("gykey", "fyear"	g values in y, and xe), "))	i all columns fro	, and y.
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > componies <- readRD select(data_object_na keeps only the variabl price_yr_end <- select filter(data_object_nau	gular r objects) IS("~/North Ameri ame, variable_n les you specify t(companies, co me, condition(s	ican Stock Market name(s)) from the datase pnm, loc, tic, fyea s))	: 1994-2018.rds") it ar, prcc_c)	Second 2: Second 2: <t< td=""><td>00000000000000000000000000000000000000</td><td>returns all merged1 < To specify f you can do merged1 <- in left_join</td><td>L rows from x inner_join(this match more the following: </td><td>where there a example1,exar e clearly (which i ,example2, by = c(' _, dataset_</td><td>mple2) is good practic "gykey", "fyear" _Y) FUNC</td><td>g values in y, and xe), ")) TION</td><td>i all columns fro</td><td>Jin X and y.</td></t<>	00000000000000000000000000000000000000	returns all merged1 < To specify f you can do merged1 <- in left_join	L rows from x inner_join(this match more the following: 	where there a example1,exar e clearly (which i ,example2, by = c(' _, dataset_	mple2) is good practic "gykey", "fyear" _Y) FUNC	g values in y, and xe), ")) TION	i all columns fro	Jin X and y.
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variabl price_yr_end <- select filter(data_object_nau finds rows/cases whe rows are kent_Otherm	gular r objects) IS("~/North Ameri ame, variable_n les you specify t(companies, co me, condition(s pre conditions sp vise they are dr	ican Stock Market name(s)) from the datase onm, loc, tic, fyea s)) pecified are true ronped	: 1994-2018.rds") it ar, prcc_c) i. If true, those	S (00001) " (000 yes) and yes) I (000 yes) " (000 yes) I (000 yes) (000 yes)	Control Prill, 1999 Control Prill, 1999 1994 1.1.1999 1994 2.1.1999 1994 2.1.1999 1994 2.1.4999 1994 2.1.4999 2004 1.1.2199 2004 1.1.4990 2004 1.1.4990 2004 1.1.4990	returns all merged1 < To specify t you can do merged1 <- in left_join return	I rows from x Is rows from x Is many from x Is many from x Is many from x Is all rows left	where there a example1,exar e clearly (which i .example2, by = c(' _, dataset_ ft_join(x, y) ret	nre matching mple2) is good practic "gvkey", "fyear" _y) FUNC ⁻ turns all row	g values in y, and xe), "') TION s from x, and all	all columns fro	and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw	gular r objects) IS("~/North Ameri ame, variable_n les you specify t(companies, co me, condition(s rre conditions sp vise, they are dr	ican Stock Market name(s)) from the datase onm, loc, tic, fyea s)) pecified are true ropped.	: 1994-2018.rds") it ar, prcc_c) t. If true, those	9 Immediation (Immediate Accord) 1 1 <td< td=""><td>0</td><td>returns all merged1 < To specify ty you can do merged1 <- in left_join return merge</td><td><pre>c- inner_join(d this match more the following: ner_join(example1 n(dataset) ns all rows lef ed2 <- left_joi</pre></td><td>where there a example1,exar e clearly (which i ,example2, by = e(' _, dataset_ t_join(x, y) ret in(example1,e</td><td>are matching mple2) is good practic ("gykey", "fyear" _y) FUNC[*] turns all row example2)</td><td>y values in y, and >>) TION s from x, and all</td><td>columns from x</td><td>and y</td></td<>	0	returns all merged1 < To specify ty you can do merged1 <- in left_join return merge	<pre>c- inner_join(d this match more the following: ner_join(example1 n(dataset) ns all rows lef ed2 <- left_joi</pre>	where there a example1,exar e clearly (which i ,example2, by = e(' _, dataset_ t_join(x, y) ret in(example1,e	are matching mple2) is good practic ("gykey", "fyear" _y) FUNC [*] turns all row example2)	y values in y, and >>) TION s from x, and all	columns from x	and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, condition sp vise, they are dr _large <- filter(co	ican Stock Market name(s)) from the datase onm, loc, tic, fyea s)) pecified are true ropped. ompanies, loc==	1994-2018.rds") at ar, prcc_c) b. If true, those "CAN", at > 1000		0 mm2 0 pric.4 0 11.0	returns all merged1 < To specify tyou can do merged1 <- in left_join return merge	<pre>c- inner_join(of this match more the following: ner_join(example) n(dataset, ns all rows lef ed2 <- left_joi</pre>	where there a example1,exar e clearly (which i ,example2, by = e(' _, dataset_ ft_join(x, y) ret in(example1,e	mple2) is good practic "gvkey", "fyear" _y) FUNC ⁻ turns all row example2)	y values in y, and);a), TION Is from x, and all	columns from x	and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variabl price_yr_end <- select filter(data_object_nau finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_l Using this function yw	gular r objects) IS("~/North Ameri ame, variable_n les you specify' t(companies, co me, conditions s vise, they are dr large <- filter(co name, new_vari large <- filter(co name, new_vari	ican Stock Market name(s)) from the datase pom, loc, tic, fyea s)) pecified are true ropped. companies, loc== iable_name=fum uitipia paw yari	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by	Image: second	(Q) (Anno	returns all merged1 < To specify 1 you can do merged1 <- in left_join return merge I	<pre>c- inner_join(« this match more the following: ner_join(example) n(dataset, ns all rows lef ed2 <- left_joi merged2 <</pre>	where there a example1,exar a clearly (which i ,example2, by = c(' _, dataset_ ft_join(x, y) ret in(example1,e <- left_j(mple2) is good practic "gvkey", "fyear" y) FUNC ⁻ turns all row example2) oin (exam	y values in y, and ee), '') TION s from x, and all ழle1, examp	columns from x	and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variabil price_yr_end <- select filter(data_object_nau finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_l Using this function, ys separating them with	gular r objects) (S("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions sp vise, they are dr large <- filter(co name, new_vari ou can create m commas.	ican Stock Market name(s)) from the datase onm, loc, tic, fyea onm, loc, tic, fyea s)) pecified are true ropped. oompanies, loc== iable_name=fun nultiple new varia	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by	Image: second	Name Participant USA Participant USA 2.000000	returns all merged1 < To specify 1 you can do you can do nerged1 <- in left_join return merge I • R	<pre>c- inner_join(« this match more the following: ner_join(example) n(dataset, ns all rows lef ed2 <- left_joi merged2 < will return J</pre>	where there a example1,exar a clearly (which i 	mple2) is good practic ("gvkey", "fyaar" _y) FUNC ⁻ turns all row example2) oin (exam	y values in y, and be), '''' TION s from x, and all uple1, examp gykey", "f	columns from x 	and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted to a b b c b c b c b c c companies of off Ex. to get average cash for each firm, enter: companies with avg ch c companies of off	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variabl price_yr_end <- select filter(data_object_nau finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_l Using this function, ys separating them with (now variable action is a select)	gular r objects) (S("-/North Ameri ame, variable_n les you specify' t(companies, co me, conditions s vise, they are dr large <- filter(cr name, new_vari ou can create m ou can create m	ican Stock Market name(s)) from the datase pom, loc, tic, fyea s)) pecified are true ropped. companies, loc== iable_name=fun nultiple new varia	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by able by) time_variable_names	(1) (returns all merged1 < To specify 1 you can do merged1 <- in left_join return merge 1 • R	<pre>c-inner_join(club) this match more the following: mer_join(club) mer_join(club) mer_join(club) the following) mer_gold <- left_join merged2 <- will return J will return J y filter</pre>	<pre>where there a example1,exar a clearly (which i</pre>	mple2) is good practic "gway", "ywar" y) FUNC" turns all row example2) oin (exam by = c ("c	y values in y, and be), ''') TION s from x, and all uple1, examp gvkey", "f:	columns from x .le2) year")	: and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(X)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted to the second	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps finds rows/cases whe rows are kept. Otherwatkeeps companies_canadian_ mutate(data_object_l Using this function, ys separating them with (new variable called fic companies %5~% muta	gular r objects) IS("~/North Ameri ame, variable_n les you specify' t(companies, co me, conditions s vise, they are dr large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 10C	ican Stock Market name(s)) from the datase pom, loc, tic, fyez s)) pecified are true ropped. iompanies, loc== iable_name=fun- nultiple new vari- is ch multiplied b 00000*ch) is ide	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by by 1 mill) entical to mutate	companies, full_ch =	Image: Non-Weight (Non-Weight (returns all merged1 < To specify i you can do merged1 < in left_join return merge I • R ch)	<pre>c- inner_join(d this match more the following: mer_join(example) n(dataset, ns all rows lef ed2 <- left_joi merged2 < will return J</pre>	<pre>where there a example1,exar a clearly (which i</pre>	re matching mple2) is good practic "www", "tynar" y) FUNC" turns all row example2) oin (exam oy = c ("(y values in y, and be), '''' TION s from x, and all uple1, examp gvkey", "f; a a a a a a a a a a a a a a a a a a a	columns from x nle2) year")	: and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted to the second	Loading rds files (sing > companies <- readRD select(data_object_na keeps only the variabl price_yr_end <- select filter(data_object_nau finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_U Using this function, ys separating them with (new variable called fic companies %>% muta nummediated variable called fic companies %>% muta	gular r objects) IS("~/North Ameri ame, variable_n les you specify' t(companies, co me, conditions s vise, they are dr large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i i tel(full_ch = 1000 com a an a	ican Stock Market name(s)) from the datase pom, loc, tic, fyea s)) pecified are true ropped. iompanies, loc== iable_name=fun- nultiple new vari- is ch multiplied b D0000*ch) is ide	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by by 1 mill) entical to mutate	companies, full_ch =		returns all merged1 < To specify i you can do merged1 < interpret left_join return merge I • R ch)	<pre>c- inner_join(d this match more the following: ner_join(example1 n(dataset, nes all rows lef ed2 <- left_joi merged2 < will return J gukey 5 fuer gukey 5 f</pre>	<pre>where there a example1,exar a clearly (which i</pre>	re matching mple2) is good practic "www", "tynar" y) FUNC" turns all row example2) oin (exam oy = c ("e	y values in y, and be), 	columns from x columns from x columns from x columns c	c and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted to be a summary of the summary of th	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_n: finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_r Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_r) used to create an azei	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, condition(s re condition(s re conditions sp vise, they are dr _large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 10C sct_name, new_ regate statistic	ican Stock Market name(s)) from the datase pom, loc, tic, fyea s)) pecified are true ropped. iompanies, loc== iable_name=fun nultiple new varii is ch multiplied t D0000*ch) is ide variable_name = over all of the o	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) putical to mutate aggregate_func bservations (or s)</pre>	Image: Second) 1000000*	returns all merged1 < To specify I you can do merged1 <- in left_join return merge I • R ch)	<pre>c- inner_join(cluster from x c- inner_join(cluster for this match more the following:</pre>	<pre>where there a example1,exar a clearly (which i</pre>	re matching mple2) is good practic "www", "tynar" y) FUNC" turns all row example2) oin (exam by = c ("(<pre>g values in y, and be), '''' TION s from x, and all pple1, examp gvkey", "f a at a a a a</pre>	columns from x sle2) year") ate a 267,660	cand y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_n Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_n) used to create an agg	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, condition(S t(companies, co me, condition(S ticompanies, co ticompanies, co tic	ican Stock Market name(s)) from the datase pom, loc, tic, fyea s)) pecified are true ropped. ioompanies, loc== iable_name=fun ultiple new vari- is ch multiplied b D0000*ch) is ide variable_name = over all of the o	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or second)</pre>	Image: Companies, full_ch = cton(existing_variable_names) 1000000*	returns all merged1 < To specify 1 you can do merged1 <- in left_join return merge I · R ch)	<pre>c- inner_join(c this match more the following: mer_join(example1 n(dataset, ns all rows lef ed2 <- left_joi merged2 </pre>	<pre>where there a example1,exar a clearly (which i ,example2, by = e(' _, dataset ft_join(x, y) ret in(example1,e <- left_jo Coining, b Coin</pre>	mple2) is good practic "gvkgy", "tyset" "gvkgy", "tyset" y) FUNC' turns all row example2) oin (example2) oin (example2) oin (example2) sis coop mc "good signal signa	gvalues in y, and be), ''''''''''''''''''''''''''''''''''''	columns from x sle2) year") ale 2 767.600 a a	cand y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_1 Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_s) used to create an agg Suppose you want to recorded for Canadia	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, condition(S tre conditions sr vise, they are dr large <- filter(co name, new_vari ou can create m commas. ull_ch which is i ate(full_ch = 100C sct_name, new_ regate statistic find the maximu	ican Stock Market name(s)) from the datase onm, loc, tic, fyea s)) pecified are true ropped. ioompanies, loc== iable_name=fun nultiple new vari- is ch multiplied t 00000*ch) is ide variable_name = over all of the o um and minimur 2010.	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash</pre>	Image: Second) 1000000*	returns all merged1 < To specify 1 you can do merged1 <- in left_join return merge I · R ch)	<pre>c- inner_join(c this match more the following: mer_join(example1 m(dataset, ns all rows lef ed2 <- left_joi merged2 < will return J gykey frear gykey frear gykey frear goldod 2015 001004 2015 001004 2016 001005 2014 5 001005 2014</pre>	<pre>where there a example1,exar a clearly (which i,example2, by = e(', dataset</pre>	re matching mple2) is good practic "gvkag", "tyast" y) FUNC turns all row example2) oin (exam by = c ("o	gvalues in y, and xe), ''''' TION s from x, and all pple1, examp gvkey", "f; a at s 1504.100 1 8500 1504.100 86300 4 86400 4 86400 4 86400 4 86500 1504.100 86400 4 86400 4 86500 51274.000	columns from x columns from x columns from x columns c	cand y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian, mutate(data_object_1 Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object) used to create an agg Suppose you want to recorded for Canadian max min cash Ca	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, condition(s ree conditions spi vise, they are dr large <- filter(co name, new_vari ou can create m commas. ull_ch which is i ate(full_ch = 100 cot_name, new_ regate statistic find the maximu in companies in unadian 2010	ican Stock Market hame(s)) from the datase brim, loc, tic, fyea s)) pecified are true ropped. iompanies, loc== iable_name=fun nultiple new vari- iis ch multiplied b 20000°ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie	1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash as \$>\$	Image: state of the state o	() 1000000*	returns all merged1 < To specify 1 you can do merged1 <- in left_join return mergy I • R ()) () () () () () () ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ()) ())) ()))) ()))) ()))) ())))) ()))))) ())))))))))))))))))))	<pre>c- inner_join(d this match more; the following: mer_join(example1 m(dataset, ns all rows lef ed2 <- left_joi merged2 < will return J grkey frear grkey frear goolood 2015 goolood 2015</pre>	where there a example1,exar a clearly (which i ,example2, by = e(' _, dataset_ ft_join(x, y) ret in(example1,e <- left_jo conm AAR CORP AAR CORP	mple2) is good practic "orkey", "tyear" y) FUNC' turns all row example2) oin (exam by = c ("o a a a sts GROUP NC 2 sts GROUP NC 2 at sts GROUP NC 2 at	y values in y, and xe), ")' TION s from x, and all mple1, examp gykey", "f; 4 1 4.000 A	columns from x columns from x columns from x columns c	c and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_1 Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object) used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_Cash_filter (loce	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions(s re conditions(s re conditions(s re conditions(s re conditions(s re conditions) large <- filter(co name, new_vari ou can create m commas. ull_ch which is i ate(full_ch = 100 cot_name, new_v regate statistic find the maximum in companies in in andian_2010 =="CAN" & fy	ican Stock Market hame(s)) from the datase born, loc, tic, fyea s)) pecified are true ropped. isompanies, loc== iable_name=fun ultiple new vari- is ch multiplied b 20000*ch) is ide variable_name = over all of the o uum and minimur 2010. 0 << companies rear==2010) %	1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash as \$>\$	the second	() 1000000*	returns all merged1 < To specify 1 you can do merged1 <- in left_join return mergy I • R • R • R • Ch)	<pre>c- inner_join(d this match more; the following: mer_join(example1 m(dataset, ns all rows lef ed2 <- left_joi merged2 < will return J going fyear 1 00004 2015 3 00104 2015 3 00104 2015 3 00104 2015 4 001045 2014 5 001045 2014 5 001045 2014 5 001045 2014 5 001045 2015 </pre>	where there a example1,exar a clearly (which i ,example2, by = e(' _, dataset_ ft_join(x, y) ret in(example1,e <- left_jo coining, b com AAR CORP AAR CORP A	mple2) is good practic "grkeg", "tyear" y) FUNC" turns all row example2) oin (exam by = c ("o is a courne" is courne" is courne" is courne" is courne" is courne" is courne" is courne" is courne" is courned in the ison provided in the ison p	y values in y, and xe), """ TION s from x, and all mple1, examp gvkey", "f; 17,700 4 at 4 at 5 10,000 4 at 4 at 5 500 10,000 4 4 at 5 600 5 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640 6 640	columns from x columns from x	(and y er
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian mutate(data_object_1 Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object) used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_Cas _filter (loc= summarise (gular r objects) IS("-/North Ameri ame, variable_m les you specify t(companies, co me, condition(s) re condition(s) re condition(s) re condition(s) re condition(s) large <- filter(cu name, new_vari ou can create m commas. ull_ch which is i ate(full_ch = 10C act_name, new_v regate statistic find the maximu n companies in n companies (s) find the maximu n companies (s) find the maximu find	ican Stock Market hame(s)) from the datase bonn, loc, tic, fyea s)) pecified are true ropped. ompanies, loc== iable_name=fun ultiple new variable is ch multiplied b D0000*ch) is ide variable_name = over all of the o uum and minimur 2010. o <- companie rear==2010) % = max (ch, na = min (ch, na	1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by by 1 mill) entical to mutate = aggregate_func bservations (or s m cash as %>% b>% .rm = TRUE), .rm = TRUE().	Image: state of the state o) 1000000*	returns all merged1 < To specify 1 you can do merged1 <- in left_join return merge I • R • R • R • Ch • Ch	<pre><- inner_join(d this match more; the following: mar_join(example1 n(dataset, ns all rows lef ed2 <- left_joi merged2 < will return J</pre>	where there a example1,exar a clearly (which i ,example2, by = c(' _, dataset ft_join(x, y) ret in(example1,e <- left_j('oining, b ' comm AAR CORP AAR CORP AAR AAR AAR AAR AAR AAR AAR AAR AAR A	re matching mple2) is good practic ("gvkag", "tywar" y) FUNC" turns all row example2) oin (exam by = c ("o s coup Nc 2 is	syndues in y, and te), ")) TION s from x, and all mple1, examp gvkey", "f; a at s b at s b at s b at s c at	columns from x heal year") aternological aterno	(and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted in the sum of the sum	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_U Using this function, yt separating them with (new variable called fi companies %>% muta summarise(data_object_U used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_Ca summarise (gular r objects) IS("-/North Ameri ame, variable_m les you specify t(companies, co me, condition(sp re	ican Stock Market hame(s)) from the datase prom, loc, tic, fyea s)) pecified are true ropped. ompanies, loc== iable_name=fum nultiple new variable is ch multiplied H 20000*ch) is ide yariable_name = over all of the o uum and minimur 2010. 0 <- companie rear==2010) % = max (ch, na = min (ch, na	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by cy1 mill) entical to mutate = aggregate_func bservations (or s m cash ess %>% .rm = TRUE), .rm = TRUE);</pre>	Image: The second se) 1000000*. (c) Non One	returns all merged1 < To specify 1 you can do merged1 < in left_join return merge I • R • R • R • R • Ch • Ch	I rows from x inner_join(d this match more the following: mer_join(example1 n(dataset) nn (dataset) nn all rows lef ed2 <- left_join merged2 will return J gold (free grad) gold (free gra	where there a example1,exar a clearly (which i ,example2, by = c(' _, dataset ft_join(x, y) ret in(example1,e <- left_j(coining, b coining, b comm AAR CORP AAR CORP AAR CORP AAR CORP AAR CORP AAR CORP AAR CORP AAR CORP AAR CORP AAR CORP AAR CO	re matching mple2) is good practic "gway", "gwar", "gwar", y) FUNC" turns all row example2) oin (exam by = c ("o a a a s coour Nc 2 st s coour Nc 2 et with no mo willant dat e matching o	s values in y, and be), ")' TION s from x, and all mple1, examp gvkey", "f; a at s base of the second a at s a at s base of the second a at s a at the second a at the sec	columns from x hee2) year") aternological toos in the other is in the therefore therefore the starter therefore the starter therefore the starter the st	and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted in the sum of the sum	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_na: finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_ Using this function, yc separating them with (new variable called fi companies %>% muta summarise(data_obje used to create an agg Suppose you want to recorded for Canadiai max_min_cash_Ca filter (loce summarise (gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions(g re conditions(g re conditions sp re conditions sp re conditions (g re conditions sp large <- filter(cu name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 100 cct_name, new_regate statistic find the maximu n companies in madian_2010 ="CAN" & fy max_cash = min_cash = lian_2010 ->	ican Stock Worket hame(s)) from the datase prom, loc, tic, fyee s)) pecified are true ropped. ompanies, loc== iable_name=fum nultiple new variable is ch multiplied H 20000*ch) is ide variable_name = over all of the o uum and minimur 2010.) <- companie = max(ch, na = min (ch, na) 1276	<pre>1994-2018.rds") it ar, prcc_c) if true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash as %>% .rm = TRUE), .rm = TRUE)) min_cash </pre>	Image: contract of the second of the seco) : 1000000*: (c) Non One When you	returns all merged1 < To specify in you can do merged1 < in left_join return merge return merge return on R ch) Obset datas "right To One M: u umerge tw	I rows from x I rows	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "gwwgr, "gwar", "gwar" y) FUNC" turns all row example2) ooin (exam oy = c ("o a s coour nc 2 as coour nc 2 et with no m soultant date we matching o	s values in y, and te,, ''' TION s from x, and all mple1, examp gvkey", "f; a at s b at s b at s b at s c a	columns from x ble2) year") at at at at at at at at at at	e and y
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [] 1 data_omit <- na.omit(data) any rows with NA are deleted in the sum of the sum o	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_na: finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_ Using this function, y separating them with (new variable called fi companies %>% muta summarise(data_object_ used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_Ca filter (loce summarise (gular r objects) IS("-/North Ameri ame, variable_m les you specify t(companies, co me, condition(s) re conditions(s) re conditions(s) re conditions(s) re conditions(s) re conditions(s) re conditions(s) re conditions(s) tice, they are dr large <- filter(cc name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 100 cc_name, new_regate statistic find the maximu n companies in madian_2010 ="CAN" & fy max_cash = min_cash = lian_2010 ->	ican Stock Market name(s)) from the datase prom, loc, tic, fyea s)) pecified are true ropped. companies, loc== iable_name=fum nultiple new variable is ch multiplied H 20000*ch) is ide variable_name = over all of the o uum and minimur 2010.) <- companie = max (ch, na = min (ch, na 1 17368	<pre>1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash as %>% .rm = TRUE), .rm = TRUE)) min_cash { 0 </pre>)) tring_variable_names companies, full_ch = companies, full_ch = companies, if specifie) 1000000* 1000000* 1000000*	returns all merged1 < To specify in you can do merged1 < in left_join return mergy r R ch) Chy Obset datas "right To One M: u merge tw ach obset datas	I rows from x inner_join(d this match more the following: mar_join(example1 n(dataset, ns all rows lef ed2 <- left_join merged2 will return J I 001004 2015 001042 2015 001042 2015 001042 2015 001042 2015 001042 2015 001045 2016 001045 2016 001045 2016 001045 2016 001045 2016 001045 2016 votataset white atching vo datasets we vation withing	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "gwwgr, "gwar", y) FUNC" turns all row example2) ooin (exam oy = c ("o a scoup wc 2 sts coop wc 2 et with no m southant dat we matching o ables unique this kind of	s from x, and all pe), """ TION s from x, and all ple1, examp gvkey", "f; a a a a s from x, and all ple1, examp gvkey", "f; a a a a s from x, and all ple1, examp gvkey", "f; a a a a a s foot a a s foot a a a a a a a a a a a a a a a a a a	columns from x le2) year")	217 A and y. (and y (and y 217 (and y 217 (and y 217 (and y 217 (and y 217) (and y (and y) (and y (and y) (and y (and y) (and y)) (and y) (and y)) (and y))) (and y))) (and y))))))))))))))))))))))))))))))))))))
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted if if i	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variable price_yr_end <- select filter(data_object_na: finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_ Using this function, yo separating them with (new variable called fi companies %-% muta summarise(data_obje used to create an agg Suppose you want to recorded for Canadiai max_min_cash_Ca filter(loce summarise(max_min_cash_Canada	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions, co re conditions, co re conditions sy re conditions sy ise, they are dr _large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 100 cc_name, new_ regate statistic find the maximu n companies in unadian_2010 ="CAN" & fy max_cash = min_cash = lian_2010 -> sks	<pre>icon Stock Worket hame(s)) from the datase prom, loc, tic, fyee s)) pecified are true ropped. companies, loc== iable_name=fun nultiple new variable_name = over all of the o uum and minimur 2010.) <- companie e max(ch, na = min (ch, na i max_cah i 1 17368</pre>	1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by cy 1 mill) entical to mutate = aggregate_func bservations (or s m cash =>% >% .rm = TRUE) , .rm = TRUE))	<pre>provide the second second</pre>	() 1000000* 1000000* 1000000* 1000000* 1000000* 1000000*	returns all merged1 < To specify ty you can do merged1 < in left_join return mergy I • R • R ch) 0 0 0 0 0 0 0 0 0 0 0 0 0	<pre>c-inner_join(c this match more the following: mer_join(example) n(dataset, ms all rows lef ed2 <- left_joi merged2 < will return J 001004 2015 001004 2015 001000 2015 000000 2015 000000 2015 000000 2015 00000000000000000000000000000000000</pre>	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "gway", "gwar" yy FUNC" turns all row example2) ooin (exam ooy = c ("d s as croup NC 2 sts croup NC 2 sts croup NC 2 sts croup NC 2 st with no m southant dat ee matching a	s values in y, and xe), '''' TION s from x, and all mple1, examp gvkey", "f; d at t s from x, and all s from x, and all mple1, examp gvkey", "f; d at t s from x, and all s from x, and all	columns from x le2) year") year") year") year") tions in the othin is in the t be returned. year downoon t and downoon and d	2) and y c and y 2) 2) 2) 2) 2) 2) 2) 2) 2) 2)
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows/cases whe rows are kept. Otherwar- companies_canadian_ mutate(data_object_U Using this function, yt separating them with (new variable called file companies %>% muta summarise(data_object used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_Ca filter (loc= summarise (max_min_cash_Canadd Common Tas: If the objective is	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions, co me, conditions, co me, conditions, re conditions sy conditions sy large <- filter(cr name, new_vari ou can create m commas. tull_ch which is i tate(full_ch = 10C sct_name, new_ regate statistic find the maximu n companies in unadian_2010 ="CAN" & fy max_cash = min_cash = lian_2010 -> sks to perform aggi	<pre>icon Stock Market hame(s)) from the datase prom, loc, tic, fyee s)) pecified are true ropped. companies, loc== iable_name=fun nultiple new variation is ch multiplied b D0000*ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie erar==2010) % = max(ch, na = min (ch, na = min (ch, na = min (ch, na) = max(ch, ch) 1 17368</pre>	1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by cy 1 mill) entical to mutate = aggregate_func bservations (or s m cash ass %>% >% .rm = TRUE) , .rm = TRUE)) min.cash 0	<pre>provide the second second</pre>	() () () () () () () () () ()	returns all merged1 < To specify ty you can do merged1 < in left_join return merge • R • R ch) 0 0 0 0 0 0 0 0 0 0 0 0 0	<pre>c-inner_join(c this match more the following: mer_join(example1 m(dataset, ms all rows lef ed2 <- left_joi merged2 < will return J</pre>	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "gway", "gwar" y) FUNC" turns all row example2) ooin (exam ooy = c ("d s as chour NC 2 is chour 2 is chour NC 2 is chour	s values in y, and xe), '''' TION s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all a from x, and all ple1 , examp gvkey", "f; d a at s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all ple1 , examp gvkey", "f; d a at s from x, and all t a at t a at	columns from x leal of the second se	2) c and y c and y 2) 2) 2) 2) 2) 2) 2) 2) 2) 2)
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows/cases whe rows are kept. Otherwa companies_canadian_ mutate(data_object_l Using this function, ys separating them with (new variable called fic companies %>% mutat summarise(data_object used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_Ca filter (loc= summarise (max_min_cash_Canadd Common Taa: If the objective is groups and to disp	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions, co me, conditions, co me, conditions, re conditions se conditions se large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 10C set_name, new_ regate statistic find the maximum, n companies in unadian_2010 ="CAN" & fy max_cash = min_cash = lian_2010 -> sks to perform aggreg but our esgrege	<pre>icon Stock Market name(s)) from the datase prom, loc, tic, fyez s)) pecified are true ropped. companies, loc== iable_name=fun nultiple new variat is ch multiplied b D0000*ch) is ide variable_name= over all of the o um and minimur 2010.) <- companie ear==2010) % = max (ch, na = min (ch, na</pre>	1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by ables by ables by ables by able to mutate = aggregate_funct bservations (or some as %>% b>% .rm = TRUE), .rm = TRUE)) min_cash 0 ables ables ab)) (companies, full_ch = ction(existing_variable subgroups, if specifie	<pre>intermediate intermediate</pre>	returns all merged1 < To specify you can do merged1 < in left_join return merge * R * R * R ch) Dise datas "right To One M: u merge tw scach obser called a other the obser bles for bosen obsers obser scach obser called a other * the attrii the obser bles for bosen bles for bosen for bles for bosen bles for bosen bles for bosen for bles for bosen for bles for bosen for bles for bosen bles for bosen for bles for	I rows from x I rows	where there a example1,exar a clearly (which i ,.example2, by = e(' _, dataset_ ft_join(x, y) ret in(example1,e c<- left_jc foining, b a comp	re matching mple2) is good practic "gway", "gwar" y) FUNC" turns all row example2) oin (exam oy = c (" s as cour NC 2 is c	s values in y, and se), """ TION s from x, and all mple1, examp gvkey", "f: 4 0 at 5 1000 a	columns from x column	2r and y and y 2r 2r 2r 2097C 41111 3009 100 3009 100 3000 1000 3000 1000 3000 3000 1000 3000 3000 3000 3000 3000 3000 3000 3000 3000 30
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(X)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows/cases whe rows are kept. Otherwatkeeps companies_canadian_ mutate(data_object_l Using this function, ys separating them with (new variable called fic companies %>% mutat summarise(data_object used to create an agg Suppose you want to recorded for Canadiaa max_min_cash_canadkeeps filter (loc= summarise (max_min_cash_Canadkeeps If the objective is groups and to diss should use group.	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions, co me, conditions, co me, conditions, re conditions si vise, they are dr _large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 10C act_name, new_ regate statistic find the maximun n companies in unadian_2010 ="CAN" & fy max_cash = min_cash = lian_2010 -> sks to perform aggreg _by() with summ	icon Stock Market name(s)) from the datase prom, loc, tic, fyez s)) pecified are true ropped. companies, loc== iable_name=fun- nultiple new varia- is ch multiplied b D0000*ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie rear==2010 % 1 17368 gate result per gr narise().	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by ables by ables by able to mutate = aggregate_funct bservations (or s m cash as %>% b% .rm = TRUE) , .rm = TRUE)) min_cash o sover roup, you)) (companies, full_ch = action(existing_variable subgroups, if specifie	<pre>intermediate intermediate</pre>	returns all merged1 < To specify you can do merged1 < in left_join return merge * R * R ch) Obser datas "right To One M: u unerge tw ach obser bles for bosen bles for bles for bles for bles fo	I rows from x I rows	where there a example1,exar a clearly (which i	re matching mple2) is good practic "gway", "gwar" y) FUNC" turns all row example2) oin (exam oy = c ("d s as as GROUPINC 7 is GROUPINC 7 is GROUPINC 7 is GROUPINC 7 is chouring a data a basis on ot have to	s values in y, and be, y, TION s from x, and all mple1, examp gvkey", "f; 4 0 at 5 0,200 64 7 150,000 1 150,100	columns from x columns from x deal dea	C and y c and
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows/cases whe rows are kept. Otherwattee companies_canadian_ mutate(data_object_J Using this function, ys separating them with (new variable called fic companies %>% mutate summarise(data_object used to create an agg Suppose you want to recorded for Canadian max_min_cash_Caa filter(loc= summarise(max_min_cash_Canadtee If the objective is groups and to dis should use group. IA The syntax is:	gular robjects) IS("-/North Americ ame, variable_n les you specify' t(companies, co me, condition(s re conditions sp vise, they are dr large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 10C sc1_mme, new_ regate statistic find the maximum in acompanies in unadian_2010 =""CAN" & fyy max_cash = min_cash = lian_2010 -> sks to perform aggreg pby() with summ data_object	icon Stock Market name(s)) from the datase prom, loc, tic, fyea s)) pecified are true ropped. companies, loc== iable_name=fun- nultiple new varia- is ch multiplied t D00000*ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie rear==2010 % in (ch, na = max (ch, na = min (ch, na = min (ch, na = min (ch, na = max (ch, na = min (ch, na = max (ch, na = min (ch, na = min (ch, na = max (ch, na = min (ch, na) = max (ch, na = min (ch, na) = min (ch, na) = min (ch, na) = max (ch, na) = min (ch, na) =	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash as %>% b>% .rm = TRUE)) min_cash o sover roup, you	companies, full_ch = chonesting_variable	<pre>></pre>	returns all merged1 < To specify i you can do merged1 < in left_join return merge * * R * R * R * R * R * R * N * * * * * * * * * * * * *	I rows from x I rows from x I nows from x I the following: I nows for a state more the following: I now following: <	where there a example1,exar a clearly (which i ,example2, by = e(' _, dataset ft_join(x, y) ret in(example1,e <- left_jd foining, b conm aaccorp accorp aaccorp accorp	re matching mple2) is good practic "gway", "gwat" y) FUNC" turns all row example2) oin (exam oy = c ("{	s values in y, and be, , , , TION s from x, and all mple1, examp gvkey", "f; 4 at s 6,500 51274.000 4 atching observation aset. Observation datasets, will no example aset. observation aset. observation ase	columns from x lease arm of the other sin the other sin the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the the returned. lease arm of the other sin the returned. lease	2) c and y c and y 2) 2) 2) 2) 2) 2) 2) 2) 2) 2)
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted iso iso iso iso iso iso iso iso iso iso	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows/cases whe rows are kept. Otherwa companies_canadian_ mutate(data_object_U Using this function, ys separating them withthe (new variable called fic companies %>% mutations summarise(data_objecture) used to create an agg Suppose you want to recorded for Canadiana max_min_cash_Caa filter(loc= summarise(max_min_cash_Canadd Common Tass If the objective is groups and to disj should use group. MA The syntax is: group_by (groute) summarise(groute) s	gular robjects) (S("-/North Americ ame, variable_n les you specify t(companies, co me, condition(s tre conditions sp vise, they are dr large <- filter(cr name, new_vari ou can create m commas. ull_ch which is i tate(full_ch = 100 cac_name, new_ regate statistic find the maximum in companies in unadian_2010 =""CAN" & fyy max_cash = min_cash = bian_2010 -> sks to perform aggreg _by() with summ data_object up_variable (variable of the section of the section of the section of the section of the section of the section of the section of the section of the section of the section of the section of the min_cash = to perform aggreg _by() with summ data_object	icon Stock Market hame(s)) from the datase prom, loc, tic, fyea s)) pecified are true ropped. companies, loc== iable_name=fun- nultiple new varia- is ch multiplied t D00000*ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie rear==2010 % in (ch, na = max (ch, sa = max (ch, sa) = max (ch, sa = max (ch, sa) = ma	1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash = s %>% >% .rm = TRUE) , .rm = TRUE)) min_cash 0 = over rooup, you	companies, fulL_ch =	Non One When you identify a summer with the	returns all merged1 < To specify i you can do merged1 < in left_join return merge 1 • R • R ch) Obser datas "right To One M: uan enge two called a or er, the attri the obser bles for bc situations, tiple obser	I rows from x I rows from x I rows from x I inter_join(d) Ithe following: Inter_join(d) Inter_join(d) Inter_join(d) Inter_ged2 Interged2 <	where there a example1,exar a clearly (which i	re matching mple2) is good practic "www", "tywe" turns all row example2) oin (exam by = c ("t s cour Nc 2 is	s values in y, and be), '''' TION s from x, and all mple1, examp gvkey", "f; a at s b b e s b b e s c of s	columns from x leal columns from x lead to the columns from x lead to the columns lead to	2) c and y c and y 2) c and y 2) 2) 2) 2) 2) 2) 2) 2) 2) 2)
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted the sum of the sum of	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_r Using this function, ys separating them withthe (new variable called fic companies %>% mutations summarise(data_objecture) used to create an agg Suppose you want to recorded for Canadiana max_min_cash_Caa filter(loc= summarise(max_min_cash_Canadd Common Tass If the objective is group_by(grouts) should use group. M The syntax is: group_by(grouts) summarise(new aggrequate furthered)	gular robjects) IS("-/North Americ ame, variable_n les you specify t(companies, co me, condition(s tre conditions sp vise, they are dr large <- filter(co name, new_vari ou can create m commas. ull_ch which is i atte(full_ch = 100 cot_name, new regate statistic find the maximum in acompanies in madian_2010 =""CAN" & fyy max_cash = min_cash = lian_2010-> sks to perform aggreg- by() with summ data_object up_variable (~ variable = cotion (exist	icon Stock Market hame(s)) from the datase prom, loc, tic, fyea s)) pecified are true ropped. companies, loc== iable_name=fun- nultiple new varia- is ch multiplied t D0000*ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie rear==2010) % is companies (ch, na = max (ch, na = min (ch, na = min (ch, na (ch, na = min (ch, s) (ch, s) = max (ch,	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate aggregate_func bservations (or s m cash ss %>% .rm = TRUE) , .rm = TRUE)) min_cash 0 ns over roup, you (s)))</pre>	companies, full_ch = ction(existing_variable subgroups, if specifie	Non One When you identify a winter When you identify a winter with multiple of the second second second second with multiple on the second with multiple on the second	returns all merged1 < To specify I you can do merged1 < in left_join return merge I • R • R ch) Obser datas "right To One M: u merge twe ach obser called a or er, the attri the obser bles for bcs situations, tiple obser or bcs returns the obser tiple obser or bcs returns tiple obser or bcs returns the obser tiple obser or bcs returns the obser tiple obser returns the obser tiple obser returns the obser tiple obser returns the obser tiple obser returns the obser tiple obser returns the obser tiple obser tiple obser returns the obser tiple obser the merget the obser tiple obser tiple obser tiple obser the merget the obser tiple obser the merget the obser tiple obser tiple obser the merget the obser tiple obser tiple obser the merget the me	I rows from x C- inner_join(c) this match more the following: mer_join(cataset, ins all rows leffer) ed2 <- left_join(cataset, ins all rows leffer) ed3 <- left_join(cataset, ins all rows leffer) revations in the retary revations in two rows into the rows in th	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "www", "www", "www", y) FUNC" turns all row example2) oin (exam by = c ("(by = c ("(b) + c))))))))))))))))))))) started by the second by t	s values in y, and be), '''' TION s from x, and all mple1, examp gvkey", "f at to atching observation atching observation atc	columns from x columns from x lease x	C and y C a
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted the sum of the sum of	Loading rds files (sing > companies <- readRD select(data_object_na finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_ru Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object used to create an agg Suppose you want to recorded for Canadian max_min_cash_Caa filter(loc= summarise(max_min_cash_Caa filter(loc= summarise) MA The syntax is: group_by(grou summarise (new aggregate_fur	gular robjects) IS("-/North Americ ame, variable_n les you specify t(companies, co me, condition(s tre conditions sp vise, they are dr large <- filter(co name, new_vari- ou can create m commas. ull_ch which is i atte(full_ch = 100 cot_name, new_vari- ou can create m commas. ull_ch which is i atte(full_ch = 100 cot_name, new_vari- regate statistic find the maximum in companies in nacdian_2010 ="CAN" & fyy- max_cash = min_cash = lian_2010 -> sks to perform aggreg- by() with summ data_object up_variable (_variable = nction (exist	ican Stock Market hame(s)) from the datase onm, loc, tic, fyea s)) pecified are true ropped. ioompanies, loc== iable_name=fun untiple new vari- is ch multiplied b 20000°ch) is ide variable_name = over all of the o um and minimur 2010.) <- companie rear==2010) % = max (ch, na = min (ch, na) = ma (c	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate aggregate_func bservations (or s m cash as %>% >% .rm = TRUE), .rm = TRUE)) min_cash 0 ms over roup, you (s)))</pre>	Image: Companies, full_ch = companies, full_ch = companies, full_ch = companies, full_ch =	Non One When you identify a variable with the second of	returns all merged1 < To specify I you can do merged1 < in left_join return merge • R • R ch) Obset datas "right To One M: u merge two sch obser called a or er, the attrif g the obser bles for bc situations, tiple obser in ematching r, the variation r, the	I rows from x I rows from x I nows from x I nows from x Inter_join (example) Inter_join (example) Inter_join (example) Inter_ged2 <- left_join Interged2 <- left_join Interged2 <- left_join Interged2 <- left_join I onoos 2014 I onoos 2014 I onoos 2014 I onoos 2015 </td <td>where there a example1,exar a clearly (which i </td> <td>re matching mple2) is good practic "www", "tywat" y) FUNC" turns all row example2) oin (exam by = c ("(constant) state of the state state of the state of the state state of the state of the state state of the state of the state of the state state of the state of the state of the state state of the state of the state of the state of the state state of the state of the state</td> <td>s values in y, and be, , , TION s from x, and all pple1, examp gvkey", "f a at s a at s at s a</td> <td>columns from x columns from x lease to the columns from x lease to the columns from x fro</td> <td>2r 2r 2r 2r 300 r 6 4000 300 r 7 40000 300 r 7 40000 300 r 7 40000 300 r 7 400</td>	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "www", "tywat" y) FUNC" turns all row example2) oin (exam by = c ("(constant) state of the state state of the state of the state state of the state of the state state of the state of the state of the state state of the state of the state of the state state of the state of the state of the state of the state state of the state	s values in y, and be, , , TION s from x, and all pple1, examp gvkey", "f a at s a at s at s a	columns from x columns from x lease to the columns from x lease to the columns from x fro	2r 2r 2r 2r 300 r 6 4000 300 r 7 40000 300 r 7 40000 300 r 7 40000 300 r 7 400
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_n: finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_r Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object used to create an agg Suppose you want to recorded for Canadia max_min_cash_Caa filter(loc= summarise(max_min_cash_Caad Common Taa: If the objective is group_and to dis should use group. A The syntax is: group_by (grou summarise (new aggregate_fur	gular robjects) IS("-/North Americ ame, variable_n les you specify t(companies, co me, condition(s; rere conditions sp vise, they are dr large <- filter(co name, new_vari- ou can create m commas. ull_ch which is i atac(full_ch = 100C sct_name, new_vari- pregate statistic find the maximum in companies in in companies in in cash = min_cash = min_cash = Sks to perform aggreg_by() with summ data_object up_variable (_variable = cction (exist	ican Stock Market hame(s)) from the datase born, loc, tic, fyee s)) pecified are true ropped. iable_name=fun ultiple new vari- is ch multiplied b D0000*ch) is ide variable_name = over all of the o num and minimur 2010.) <- companie rear==2010) % = max (ch, na = min (ch, na) = m	<pre>1994-2018.rds") it ar, prcc_c) . If true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate aggregate_func bservations (or s m cash as %>% .rm = TRUE), .rm = TRUE)) min_cash s over roup, you (s)))</pre>	ecompanies, full_ch	Non One When you identify y with multi one-to-oo . However should ut	returns all merged1 < To specify you can do merged1 < in left_join return merge • R • R • R ch) Obserding to the second of the second sec	I rows from x I rows from x I rows from x I match more	where there a example1,exar a clearly (which i ,example2, by = c(' _, dataset_ ft_join(x, y) ret in(example1,e <- left_j('oining, b comm AAR CORP AAR CORP	re matching mple2) is good practic "www", "tywar", y) FUNC" turns all row example2) oin (exam by = c ("(c stantion of the stanting state with no matching ables unique this kind of ed as a basis o not have to asset can mat test through r m the merg rtasets.	s values in y, and be, y, TION s from x, and all all all all all bloc bloc c c c c c c c c c c c c c	columns from x columns from x clear columns columns clear columns clear columns clear	2 c and y c and y 2 c and y 2 2 2 2 2 2 2 2 2 2 2 2 2
<pre>evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted in the sum of the s</pre>	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variabl price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_1 Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object used to create an aggi Suppose you want to recorded for Canadia max_min_cash_Ca filter(loc= summarise(max_min_cash_cand Common Tas: If the objective is group_s and to dis should use group. IA The syntax is: group_by(grou summarise (new aggregate_fur main data types in Exce	gular robjects) IS("-/North Americ ame, variable_n les you specify t(companies, co me, conditions, co me, conditions, co me, conditions, co me, conditions, co me, conditions, co me, conditions, co aname, new_vari- ou can create m commas. Ul_ch which is i atac(full_ch = 100 cot_name, new_vari- ou can create m commas. Ul_ch which is i atac(full_ch = 100 cot_name, new_vari- gate statistic find the maximum in companies in in companies in in cash = min_cash = min_cash = sks to perform aggreg- play one aggreg_ play one aggreg_ play one aggreg_ play one aggreg- play one a	ican Stock Market hame(s)) from the datase born, loc, tic, fyee s)) pecified are true ropped. ompanies, loc== iable_name=fun ultiple new variable is ch multiplied b D0000*ch) is ide variable_name = over all of the o uum and minimur 2010.) <- companie rear==2010) % = max (ch, na = min (ch, na (nax cash = 1 17368 pregate operation gregate operation marise(). \$>\$ 	<pre>1994-2018.rds") it ar, prcc_c) i. If true, those "CAN", at > 1000 ction_name(exis ables by cy1 mill) entical to mutate = aggregate_func bservations (or s m cash >>% .rm = TRUE), .rm = TRUE)) ns over roup, you (s)))</pre>	(companies, full_ch = companies, if specifie	Non One When you identify e merge is . However matching key varia . However should ut	returns all merged1 < To specify i you can do merged1 < in left_join return merge • R • R • R ch) Obserd datas "right To One M: u merge two sach obserd situations, tiple obser ne matching er, the attrift g the obser in ematching er, the varian in quely ide wo datase	I rows from x Inter_join (example) Inter_join (example) Inter_ged2 < I rows left I control (example) I control (ex	where there a example1,exar a clearly (which i ,axample2, by = c(' _, dataset_ ft_join(x, y) ret in(example1,e <- left_j('oining, b comm AAR CORP AAR CORP	re matching mple2) is good practic ''gway'', "gwar'', "gwar'', y) FUNC' turns all row example2) oin (exam by = c ("o a a a s chour Nc 2 as ch	s values in y, and be, y, TION s from x, and all pple1, examp gvkey", "f at at s at s a	columns from x columns from x from from x from x fro	2 c and y c and y 2 c and y 2 c and y 2 2 2 2 2 2 2 2 2 2 2 2 2
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted in the answer of the a	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variable price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian_ mutate(data_object_r) Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_r) Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_r) Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_r) Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_object_r) max_min_cash_Canad Common Tax If the objective is group_by (grou summarise (new aggregate_fur hain data types in Exce E or FALSE hole numbers (e.g., 1.	gular r objects) IS("-/North Ameri ame, variable_n les you specify t(companies, co me, conditions; re conditions; re conditions (se re conditions (se re conditions) ull ch which is i ate(full_ch = 100 commas. ull ch which is i ate(full_ch = 100 commas. ate(full_ch = 100 commas. a	ican Stock Market name(s)) from the datase onm, loc, tic, fyea s)) pecified are true ropped. ompanies, loc== iable_name=fum is ch multiplied H 20000*ch) is ide variable_name = over all of the o uur and minimur 2010. ot <= companie rear==2010) % = max (ch, na = min (ch, na (max_cash = 1 17368 regate operatior gate result per gran marise(). \$>% :)) \$>% : :ing_variable	<pre>1994-2018.rds") it ar, prcc_c) if true, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash as %>% .rm = TRUE), .rm = TRUE)) min_cash 0 min_cash () </pre>	You want to make the follow water follow	Non One When you identify e merge is should un energe the twing variab with multione-to-or should un	returns all mergedl < To specify i you can do mergedl < in left_join return merge * R * R * R * R * R * R * R * R	I rows from x Intermediate rows left I constant	where there a example1,exar a clearly (which i ,example2, by = c(' _, dataset_ ft_join(x, y) ret in(example1,e <- left_j(foining, b comm AAR CORP AAR CORP	re matching mple2) is good practic is good practic is good practic is good practic is good practic turns all row example2) op = c ("o a to complete as choop let c as c as choop let c as c as choop let c as c as choop let c as c as c as c as c as c as c as c as	s values in y, and be), '''' TION s from x, and all mple1, examp gvkey", "f; a a a a s from x, and all mple1, examp gvkey", "f; a a a a s from x, and all mple1, examp gvkey", "f; a a a a a s from x, and all mple1, examp gvkey", "f; a a a a a s from x, and all mple1, examp gvkey", "f; a a a a a a s from x, and all a	columns from x clear () columns from x clear () columns from x clear () cle	2r and y and y 2r 2r 2r 2017 2010 2019 2
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(is.na(x)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted in the companies of observations based on one or more variables Ex. to get average cash for each firm, enter: companies_with_avg_ch <- companies %>% group_by(gvkey) %>% mutate(cash_avg = mean(ch, na.rm = TRUE)) if the companies_with_avg_ch <- companies %>% group_by(gvkey) %>% mutate(cash_avg = mean(ch, na.rm = TRUE)) if the companies to the console if the companies the console if the companies to the console if the companies t	Loading rds files (sing > companies <- readRD select(data_object_n: keeps only the variable price_yr_end <- select filter(data_object_nat finds rows/cases whe rows are kept. Otherw companies_canadian, mutate(data_object_U Using this function, ys separating them with (new variable called fi companies %>% muta summarise(data_objeu used to create an agg Suppose you want to recorded for Canadia max_min_cash_Canad filter (loc= summarise (max_min_cash_Canad Common Tas: If the objective is groups and to dis should use group. A The syntax is: group_by (grou summarise (new aggregate_fur hain data types in Excee E of FALSE hole numbers (e.g., 1, - b, or -55.53).	gular r objects) S("-/North Ameri ame, variable_m les you specify t(companies, co me, conditions; re conditions; re conditions (sp re conditions (sp re conditions) ull_ch which is i tate(full_ch = 100 commas. ull_ch which is i tate(full_ch = 100 commas. state(full_ch = 100 commas. state(full_c	<pre>icon Stock Worket hame(s)) from the datase prom, loc, tic, fyee s)) pecified are true ropped. oompanies, loc== iable_name=fum ultiple new variable is ch multiplied H D0000*ch) is ide variable_name = over all of the o uum and minimur 2010.) <- companie ear==2010) % = max(ch, na = min (ch, n</pre>	<pre>1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate = aggregate_func bservations (or s m cash >>% .rm = TRUE), .rm = TRUE)) min_cash * 0 min_ca</pre>	you want to m have the follow	Non One When you identify y merge is · However matching key varial · However should un rerge the twing varials cription	returns all merged1 < To specify in you can do merged1 < in left_join return merget • R • R ch) Ch) Ch) Ch) Ch) Ch) Ch) Ch) C	I rows from x I rows	where there a example1,exar a clearly (which i ,example2, by = c(: , dataset	re matching mple2) is good practic "gway", "gwar" yy) FUNC" turns all row example2) oin (exam oy = c ("c a to complexe sts chour NC 2 as chour NC 2 as chour NC 2 as chour NC 2 at with no m sublath data e matching a ables unique this kind of as a basis o not have to aset can mat ats through r m the mergitasets. ation will h, and lustry	s values in y, and xe), '''' TION s from x, and all mple1, examp gvkey", "fj a at s monormality of a second second second second a second s	columns from x clear to the cle	21 A and y.
evaluate to 1, thus giving the number of missing elements. is.na() also works with data frames. > sum(1s.no(X)) [1] 1 data_omit <- na.omit(data) any rows with NA are deleted if if i	Loading rds files (sing > companies <- readRD select(data_object_natkeeps only the variable price_yr_end <- select filter(data_object_natkeeps only the variable rows are kept. Otherwa companies_canadian_ mutate(data_object_l Using this function, ys separating them with (new variable called fic companies %>% mutat summarise(data_object used to create an agg Suppose you want to recorded for Canadian max_min_cash_Canadian max_min_cash_Canadian Common Tas: If the objective is groups and to dis sfour by (group summarise (new aggregate_fur the objective is sound to dis sound to	gular r objects) IS("-/North Ameri ame, variable_m les you specify t(companies, co me, condition(s) re conditions(s) re conditions(s) re conditions(s) re condition(s) re conditions(s) re condition(s) re condition(s) tice(full_ch = 100 cot_name, new_r regate statistic find the maximu n companies in madian_2010 ="CAN" & fyy max_cash = min_cash = min_cash = lian_2010 -> sks to perform aggreg _by() with sumn data_object up_variable = nction (exist all -13, or 678) or ich cannot be re	<pre>icon Stock Worket hame(s)) from the datase prome to the datase proped. from the datase proped. from the datase proped. from the datase fr</pre>	<pre>1994-2018.rds") it ar, prcc_c) it frue, those "CAN", at > 1000 ction_name(exis ables by oy 1 mill) entical to mutate aggregate_func bservations (or s m cash ass %>% >% .rm = TRUE), .rm = TRUE); min_cash 0 ms over roup, you (s))) rs ber or</pre>	You want to m have the follow NAICS_desc description to	Non One When you identify e merge is • Howeve matching key varial • In such with multione-to-o • Howeve should un erege the the virg variabuse science of the should un every obset	returns all merged1 < To specify in you can do merged1 < in left_join return merget • R ch) Chi Chi Chi Chi Chi Chi Chi Chi	I rows from x I rows	where there a example1,exar a clearly (which i 	re matching mple2) is good practic "gway", "gwar" y) FUNC turns all row example2) oin (exam by = c ("(1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	s values in y, and se), """ TION s from x, and all ple1 , examp gvkey", "f;	columns from x lease to the second s	21 and y 21 and y

Text Type (string of characters): Data which cannot be resolved as number or logical type is stored as text type (up to 32,768 characters). All characters can be stored as text ($h=2, a=z, 0=9, 1 \notin \$$ etc.) Error Type: Values returned when Excel encounters an error in evaluating the

Error Type: Values returned when Excel encounters an error in evaluating the contents of a cell (#DIV/0, #N/A, #NAME?, #NULL!, #NUM!, #REF!, and #VALUE!)

merged3 <- left_join(example3, NAICS_2_6_digit_codes, by = c("naicsh" = "NAICS"))</pre>

What	Syntax	Description			Example		 Some comr 	non logical operators:	
IF	=IF(logical_test, [value_if_true], [value_if_false])	Logical Operator		=IF(B2>D2, "Goo	od", "Bad")		=	Equal to	
MAX	=MAX(number1 [number2])			A	8 C		<>	Not equal to	
WIFU'		values can be numbers (e.g. 10), a cell reference that contain numbers (e.g. A8), a cell range/array that contains numbers (.S e.g.	1 Data 2 69 3 76			<	Less than	
MIN	=MIN(number1, [number2],)	A8:A10), or a combination of the three.		4 80 5 87 6 91			>	Greater than	
SUM	=SUM(number1, [number2],)			7 Formula W 8 =MAX(A2:A5) Returns the 9 =MIN(A2:A6) Returns the	At the Formula Does Result e largest value in range A2:A6. 91 e smalles value in range A2:A6. 69		<=	Less than or equal to	
AND	=AND(logical1, [logical2],)	Used to evaluate if all logicals are TRUE or FALSE		10 SOM(A2A6) Adds up the	e values in range ACNO. 403		~=	Greater than or equal to	
OR	=OR(logical1, [logical2],)	Used to evaluate if at least one of the logicals are TRUE or FA	LSE						
COUNTIES									
COUNTIES	=COUNTIFS(criteria_range1, criteria1,)	Count values in a specific range that meet certain criteria. At least 1 criterion is needed, but up to 127 criteria can be tested. Criteria range refers to range of cells you want to	test for	Ex. How many NY airports had at least 10 different airlines and at least 1 or million passenger departures in 2014?					
		Criteria can be in the form of a number, cell reference, or text. Do NOT use formulas/functions for criteria		1) airlines >=10 2) passenger de	partures >=1000000				
				=COUNTIFS(D3:	:D7,">=10",C3:C7,">=100	0000")			
				returns amount	that fulfills the criteria.				
SUMIFS	=SUMIFS(sum_range,	Adds values in a specified range or		Fx What are the	e total passenger departi	ires at NY airn	orts that had at		
	criteria_range1, criteria1,)	multiple ranges that meet specified criteria. At least 1 criterion is needed, but up to 127 criteria can be tested. Same	e as	Ex. What are the total passenger departures at NY airports that had at least 10 different airlines and at least 1 million passenger departures in 2014?					
	criteria_range: the values you're checking criteria1: what do you want it to equal to?	COUNTIES, but sums it up arterward.		Our sum_range	is passeriger departures	(03.07).			
		IMPORTANT: your sum_range must be of the same size and shape with your criteria_range. Ex.		=SUMIFS(C3:C7,D3:D7,">=10",C3:C7,">=1000000")					
		can use D3:D7 as your criteria range because both are		1 Order ID Product Unit Price Quantity =VLOOKUP(10251,A1:B6 2,					
	=VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])	of the same size (5 cells going downward), but you cannot use D3:D8 as your criteria range.		2 10247 3 10249 4 10250	Apples \$14.00 Oranges \$9.80 Bananas \$34.80	12 10 5			
VLOOKUP	=VLOOKUP(VALUE, table, index_number, approx_match)	The VLOOKUP function looks up a value in the first column o	fa	5 10251	Pears \$18.60				
	value The value to look for in the first column of the table.	range of cells, and then returns a value from any cell on the s row of the range.	ame	6 10252 7 column	Grapes Star 1 column 2	40			
	table: Two or more columns of data that's sorted in ascending order	exact matches, or approximate matches: the nearest value the than or equal to the search value.	ıat is less	Ess F1 I X J s VLOOKUP(10251,A1:86,2,FALSE) A B C D E F G H J 1 Order ID Product Unit Price Quantity Pears J					
	index_number: The column number in the table from which the matching value must be returned. The first number is one.	For example, you can use VLOOKUP to assign a grade of B+ for students who receive a grade from 76 to 79.		2 10247 Ap 3 10249 Or 4 10250 Bai 5 10251 Pe 6 10252 Grave	spies \$14.00 12 ranges \$9.80 10 inanas \$34.80 5 rars \$18.60 9 rapes \$42.30 40	~			
	approximate_match:	If the VLOOKUP is used to find an exact match, the rows		8					
	FALSE = EXACT MATCH,	can be in any order.		Approximate Mat	tches		90 - 100 A+ 85 - 89 A	0 F 50 D	
	TRUE = APPROXIMATE MATCH	VLOOKUP starts with the first value in the first column,		If you want to lo	ok up a value in a range (e.g	76 to 79),	80	55 C-	
	(if omitted, TRUE is DEFAULT	and works its way down until it finds a match.		sorted from the	nge the lookup table so that to lowest to the highest.	le data are	72 - 75 B	64 C+	
		Therefore, to avoid mistakes, make sure that each value		• You MUST only	y include the lowest value i	the range	68 - 71 B- 64 - 67 C+	68 B- 72 B	
		duplicates in the first column.		(in this case, 76). It is called the breakpoint.		60 - 55 - 59	76 B+ 80 A-	
				 Do NOT use the 76 – 79) 	e complete range in Column	(i.e. NOT	50 - 54 D	85 A 90 A+	
I FFT	=I FFT(fext num chars)	LEET function raturns the first character(s) in a string of		Faculty is = RIG Status is =LEFT Year Entered is	GHT(A2,1) F(A2,1) s =MID(A2,5,2	'	NCORRECT lookup table	CORRECT lookup table	
RIGHT	=RIGHT(text.num chars)	text based on the specified number of characters.	ſ		B (2).) OR =C2&"	"&B2	F G	
		return the last character(s) in a string of text based on the sp number of characters.	ecified	1 Student	No LastName First A Hadfield Chris	ame Status D	S Year Entered Fa	Aculty Full Name Chris Hadfield	
MID	=MID(text,start_num,num_chars)	MID function returns character(s) in a string of text,	ī	3 D38512C Gretzky Wayne D 12 C Wayne G 4 I49511A Winfrey Oprah I 11 A Oprah Wi					
	start_num: the (one indexed, inclusive) number to start at.	starting at a specified position and based on the specified number of characters.		5 I55313A	Lin Jeren	iy I	13 A	Jeremy Lin	
	num_cnars: the number of characters to include.		-1	6 I52312C	Clinton Hillar		12 C	Hillary Clinton	
CONCATENATE	=CONCATENATE(text1,[text2],) =C2&" "&B2 (alternate)	CONCATENATE function join two or more strings of text into one. Up to 255 strings of text can be joined.	1 Co	urse Code/No	Short Course Name	Char Count	t Trimmed Course	Code BUSI Course Code	
		Instead of using the CONCATENATE function, we could		MM 293	MANAGERIAL ACCTNG	10	6 COMM 293	BUSI 293	
		also use the ampersand (&) operator to join multiple strings of text into one	4 CO	MM 329	ORGANIZTL BHVR	14	4 COMM 329	BUSI 329	
	-I EN/(text)		5 CO	MM 335	INFO SYS MNGT	13	3 COMM 335	BUSI 335	
		string. Spaces are counted, too.	Trimmed	Course Code is	=TRIM(A2)	1 1		0031333	
TRIM	=TRIM(text)	TRIM function removes all spaces from text strings,	Char Cou BUSI Cou	unt is =LEN(A2) urse Code is =SU	JBSTITUTE(D2,"COMM",'	BUSI").	New Course Cod	le is =REPLACE(A9,6,1,2)	
CURCTITUTE		EXCEPT for single spaces between words.	G Old Cour	se Code Form	H I	Code	A Old Course G	B ade New Course Code	
SUBSTITUTE	m])	Substitutes a new text string for an old text string.	COMM 3	35 =SUBSTIT	TUTE(G2,3,2,1) COMM 235	code	9 COMM 335	COMM 235	
	Instance_num: optional, If you use instance number			35 =SUBSTIT 35 =SUBSTIT	TUTE(G3,3,2,2) COMM 325 TUTE(G4.3.2) COMM 225		10 COMM 396	COMM 296 COMM 297	
	the old_text in that specific instance.	Ex. Dean wants to change COMM 335 to COMM 235. If you decide to use SUBSTITUTE function	4 4	A B	C		D		
	REPLACE function replaces an old text string with a	to change the course number, instance_number is	1 Within 2 Missis	n Text Formula Us sippi =FIND("issi",A	sed Result A2) 2 Starts from the	irst character ("M	Description A") and returns the starting	position of the first "issi" found.	
	new text string, starting from a specific location of text	the first time Excel finds the number 3. The instance	3 Missis	sippi =FIND("issi",A	A3,3) 5 Starts from the Besults in #VAL	hird character (th	e first "s") and returns the	starting position of the first "issi" for characters of "ISSI" in the text string	
REPLACE	string and based on a specified number of characters. =REPLACE(old text start num num chars new text)	number is, therefore, 1.	4 Missis	sippi =FIND("ISSI",4	A4) #VALUEI Recall that FIND	is case sensitive.			
	,,			A	В		С		
FIND	=FIND(find_text,within_text,[start_num]) Find text: refers to the text you want to find. If find text	The FIND function does NOT allow wildcard characters.	1 Wildcar	d Char Description	Example Use n??n to search	or any text that has	any two characters between	the first and second n. Examples: noun or	
	is not in the within_text, the FIND function will return the	The FIND function is case sensitive.	2 ?	Any single character. One of a second n. You search for non, however, it won't work since there is only one character in between the first and second n. You search for any text that has any series of characters beginning with n and ending with n. Examples: a Any number of characters. You search for any text that has any series of characters beginning with n and ending with n. Examples: nour, noon, non, mining, business in China, and even just m(tero is also considered a valid number). Software of the search for any text that has any series of characters beginning with n and ending with n. Examples: nour, noon, non, mining, business in China, and even just m(tero is also considered a valid number). Software of the search for any text that has any series of characters beginning with n and ending with n. Examples: To look for * asterisk. Sample: To look for * asterisk. Example: To look for * stater: Example: To look for * tide. Example: To look for * tide. Example: To look for * tide. The search for any text that the search for any text					
	erter #VALUE:		3						
	Start_num: is optional. Start_num refers to the character	The SEADCH function - II	4 ~ follow 5 ~ follow						
	text.	characters and is NOT case sensitive.	6 ~ follow						
SEARCH	=SEARCH(find_text,within_text,[start_num])		1 Within 2 Mississ	Text Form	2) Result 2	s from the first chara	Description acter ("M") and returns the star	ting position of the first "issi" found.	
Wildoard of	paracters		3 Mississi	ppi =SEARCH("issi",A3	3,3) 5 Start Start	from the third chara from the first chara	acter (the first "s") and returns acter ("M") and returns the star	the starting position of the first "issi" found ting position of a series of characters that	
wildcard Cr		=MID(A2,FIND(",",A2)+2,2).	4	-32Anton(111,A4)	begi Start	s with "i" and ends w s from the first chara	with "i" with any number of cha acter ("M") and returns the star	racters in between. ting position of a series of characters that	
• ? is u	sed to find any single character. One ? is	A B 1 City, Province/State, Country Province/State 2 Auctie TX lighted States	5 Mississi	ippi =SEARCH("i??i",AS	5) 2 begi in be	s with "i" and ends w tween).	with "i" with 2 characters in be	ween (number of ? = number of characters	
equal	to one character.	3 Edmonton, AB, Canada AB Halifax, NS, Canada NS	6 Mississi	search("i*i",A6,	,6) 8 Start char	from the sixth chara acters that begins wit	acter (the third "s") and returns th "i" and ends with "i" with an	the starting position of a series of y number of characters in between.	
• is us	sod to find an actual 2 * or whether that	5 Kansas City, MO, United States 6 San Francisco, CA, United States CA	Mississi	ppi =SEARCH("issi",A7	7,SEARCH("issi",A7)+1) 5 SEAR	s similarly to the second CH("issi",A7)+1 returns	ond example above. Starts from irns the number 3. Returns the	a the third character since the start_num starting position of the first "issi" found	
 ~ IS U string 	set to find an actual r , , or ~ inside a text	7 Seattle, WA, United States WA 8 Toronto, ON, Canada ON 9 Vancuer, BC, Canada Br	8 Mississi	ppi =SEARCH("i?i",A8)) #VALUEI Resu	the third character (t ts in #VALUE! error v	value because there is no series	of characters that begins with "i" and end	
Jung					With				