Getting more out of dplyr

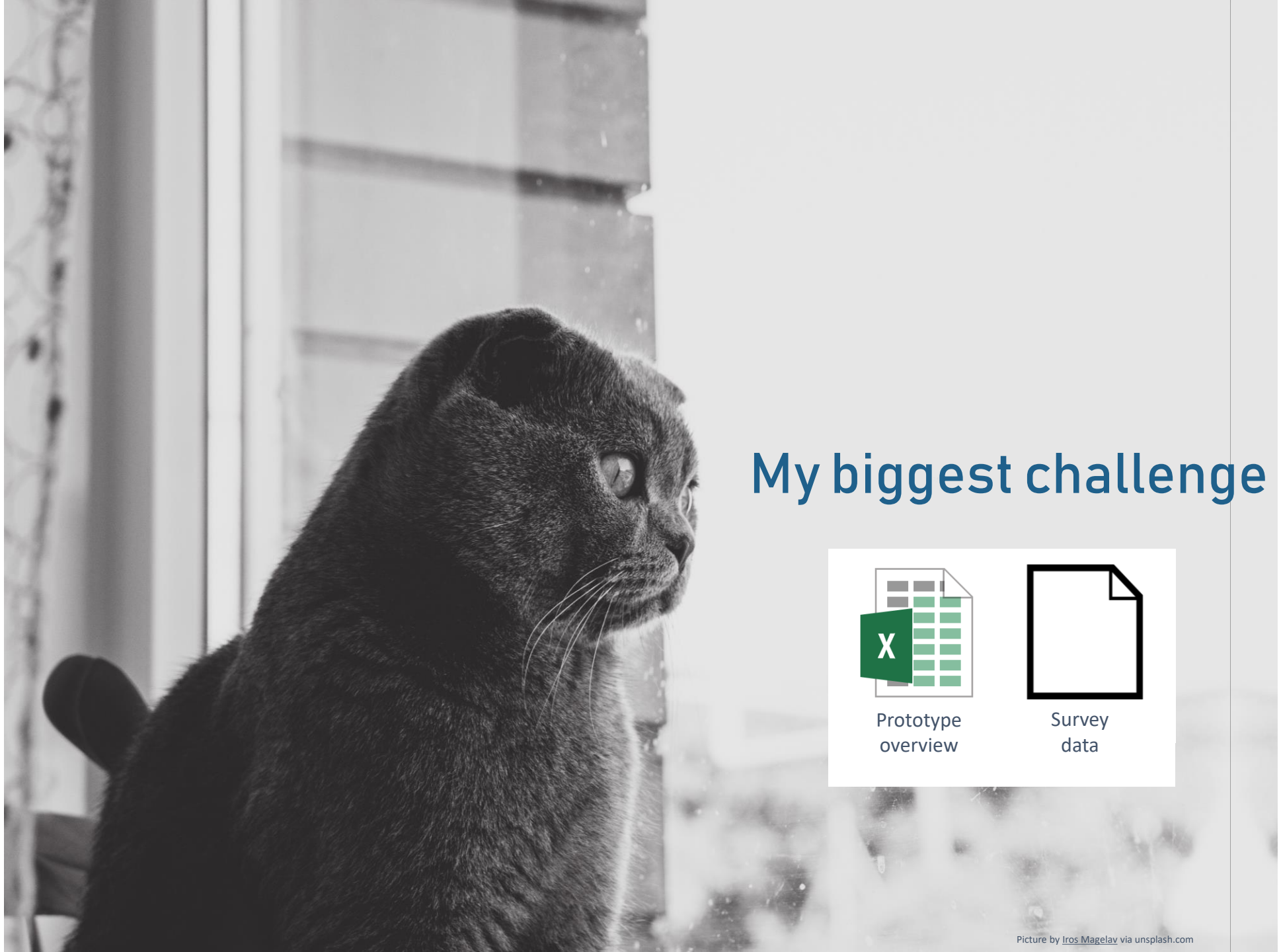
Suzan Baert | @suzanbaert
BI and analytics consultant at Hedera

SatRdays Amsterdam 2018

my journey

my data analysis journey

Picture by Johannes Plenio via pexels.com



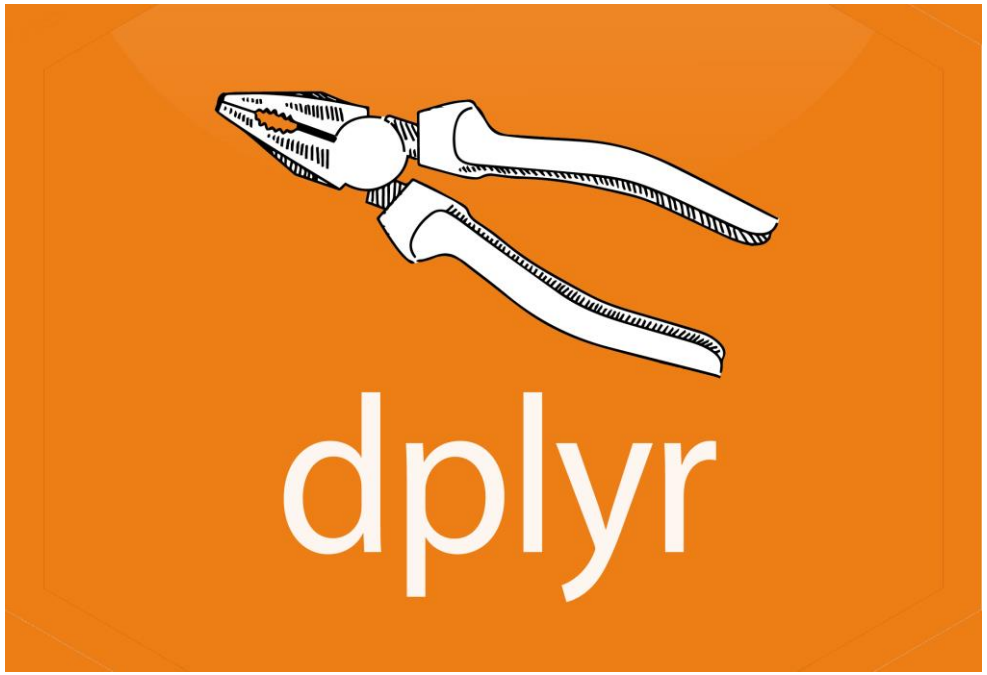
My biggest challenge

Prototype overview

Survey data

Picture by Iros Magelav via unsplash.com

My multi column challenge



Package made and maintained

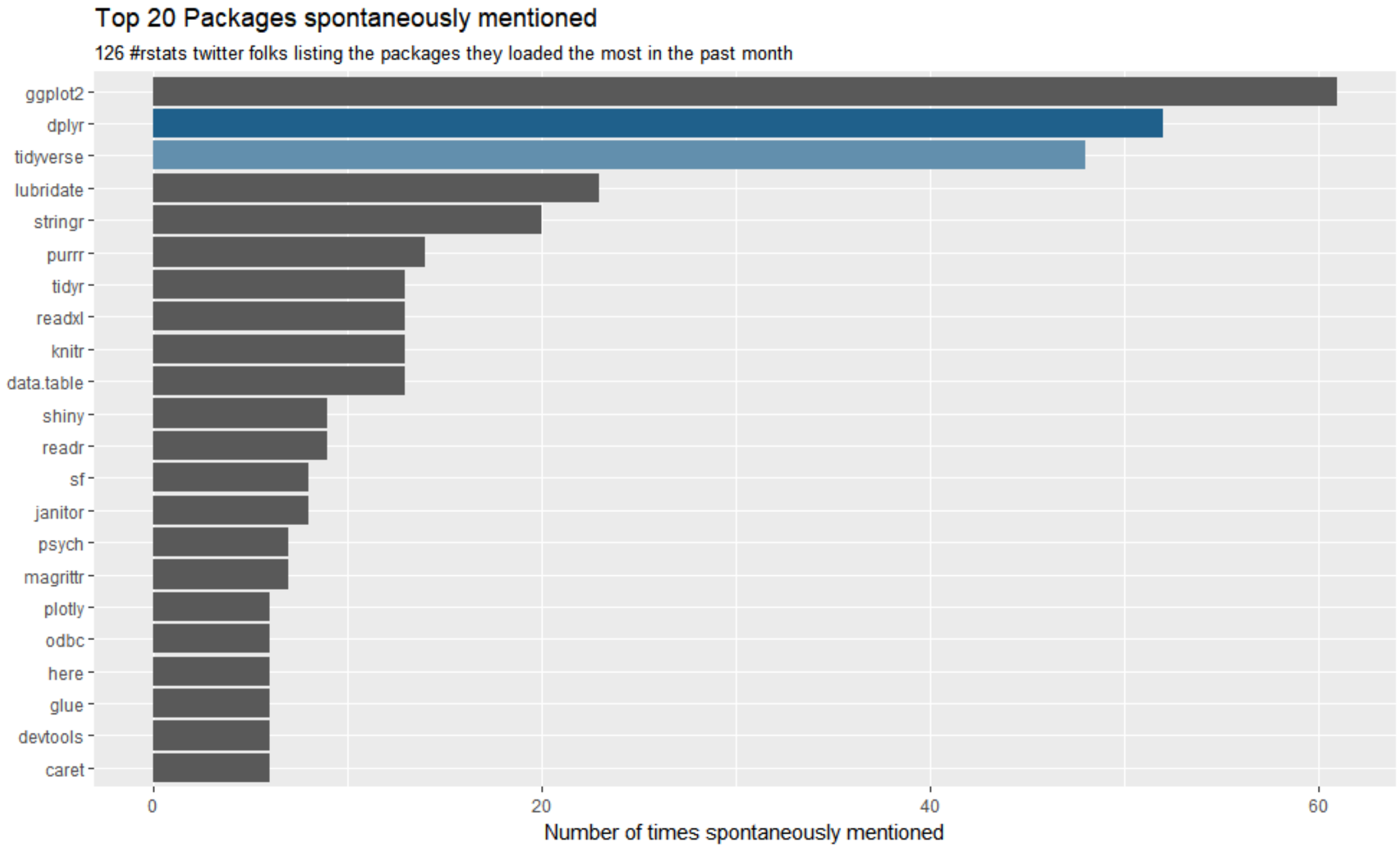
- Hadley Wickham
- Romain François
- Lionel Henry
- Kirill Müller
- other R studio folks

Today's sample data set

Warning: Not a serious or decent survey, and not intending to be.
Danger: Huge selection bias at play. Not representative of R community.

Thinking about the last month or so, what are the packages you loaded the most when using R?

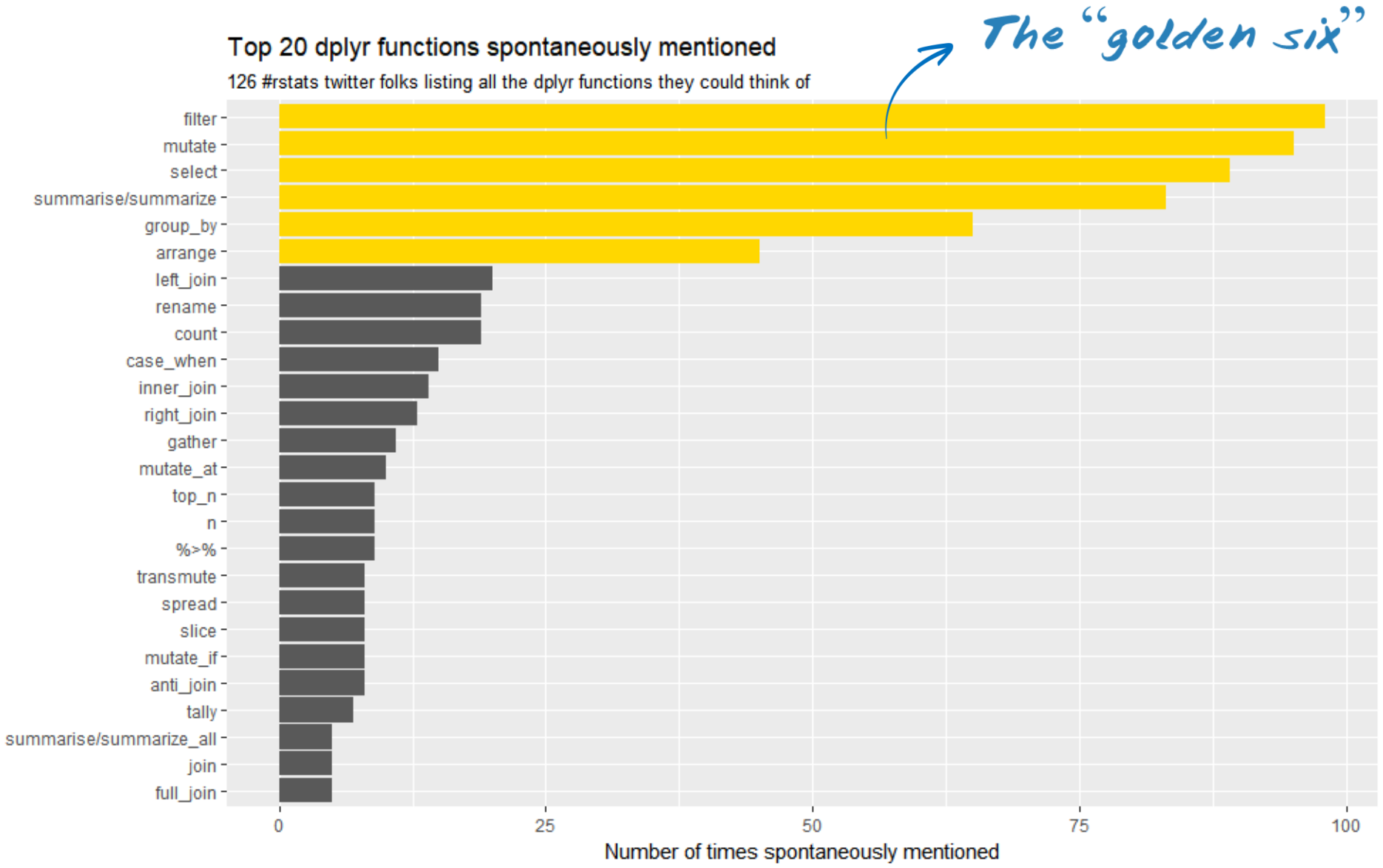
Your answer


Top 20 Packages spontaneously mentioned
126 #rstats twitter folks listing the packages they loaded the most in the past month
ggplot2
dplyr
tidyverse
lubridate
stringr
purrr
tidyr
readxl
knitr
data.table
shiny
readr
sf
janitor
psych
magrittr
plotly
odbc
here
glue
devtools
caret
0
20
40
60
Number of times spontaneously mentioned

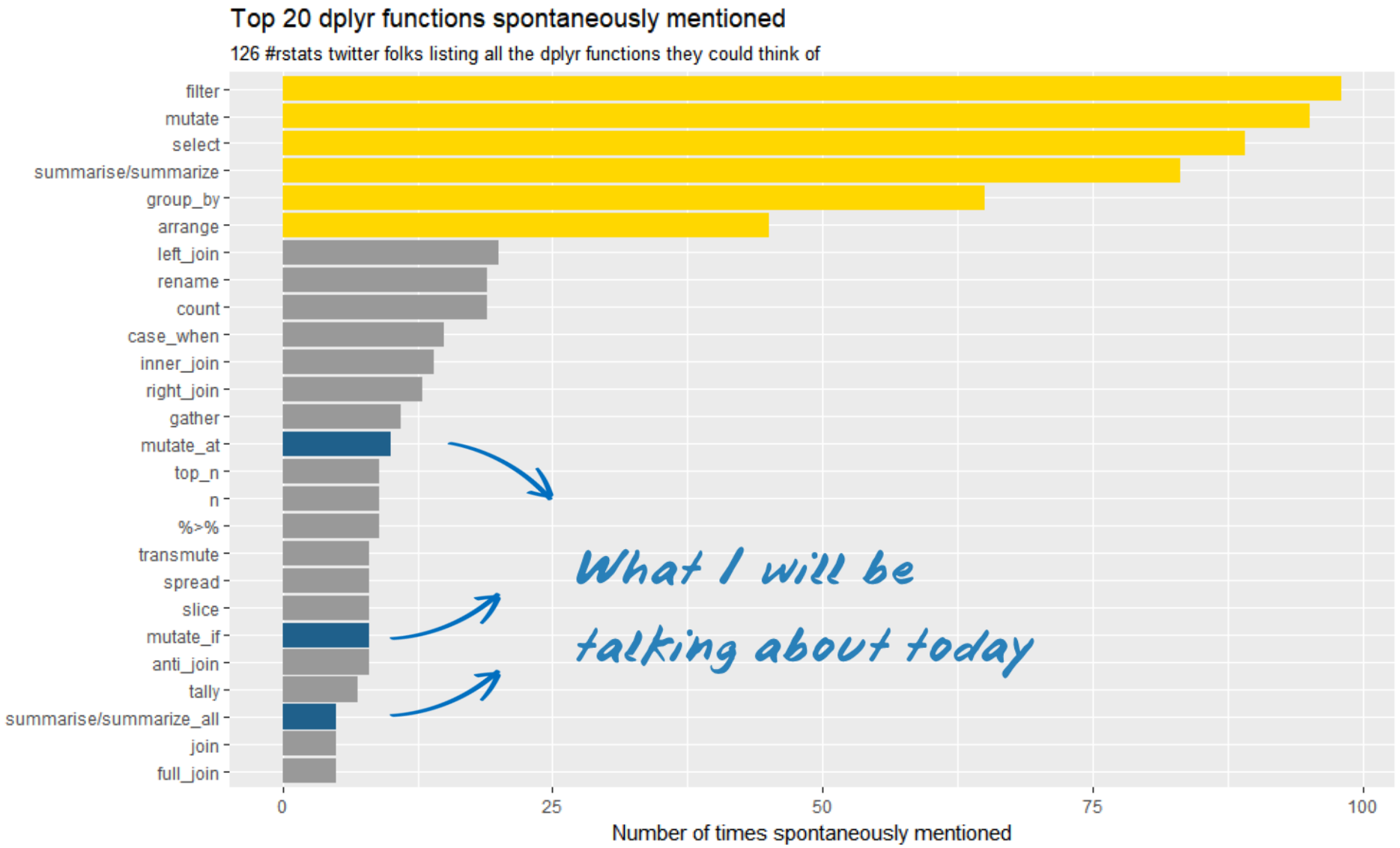
Specifically about dplyr: can you list below all dplyr functions you can spontaneously think of right now?

Your answer

dplyr basics



dplyr basics



I've taken a set of frequently downloaded packages
logs on Harry Potter's most recent birthday). From th
set, how often do you load functions from these pac

	Always	Usually	About half the time	Seldom
Rcpp	○	○	○	○
stringi	○	○	○	○
ggplot2	○	○	○	○
glue	○	○	○	○
rlang	○	○	○	○
stringr	○	○	○	○
dplyr	○	○	○	○
pillar	○	○	○	○
fansi	○	○	○	○
tibble	○	○	○	○
utf8	○	○	○	○
jsonlite	○	○	○	○
curl	○	○	○	○
purrr	○	○	○	○
data.table	○	○	○	○

Below are some quotes taken from brainyquote.com or from my private collection. Can you indicate how much you agree or disagree with each quote?

	Strongly agree	Slightly agree	Agree nor disagree	Slighly disagree	Strongly disagree
"Do or not do. There is no try." (Yoda)	○	○	○	○	○
"It does not matter what people call you unless they call you pigeon pie and eat you up." (Evelyn Waugh)	○	○	○	○	○
"Get your facts first, then you can distort them as you please." (Mark Twain)	○	○	○	○	○
"You should laugh every moment you live, for you'll find it decidedly difficult afterwards." (Joe Abercrombie)	○	○	○	○	○
"I love deadlines. I love the whooshing noise they make as they go by." (Douglas Adams)	○	○	○	○	○
"Behind every great man is a woman rolling her eyes." (Jim Carrey)	○	○	○	○	○
"Winter is coming." (George RR Martin)	○	○	○	○	○

Lastly, to get some demographics:

How many years have you been using R? Please enter the number of years below. If less than a year, you can use decimals (e.g. 0.5 for months)

Your answer

In which continent do you currently live?

- ○ North America
- ○ South America
- ○ Europe
- ○ Africa
- ○ Asia
- ○ Oceania
- ○ Antarctica

Which generation do you belong to?

- ○ Silent Generation (born before 1942)
- ○ Baby Boomers (born approx. 1943-1960)
- ○ Generation X (born approx. 1961-1982)
- ○ Millenials (born approx. 1983-2000)
- ○ Generation Z (born approx. after 2000)
- ○ None of your business

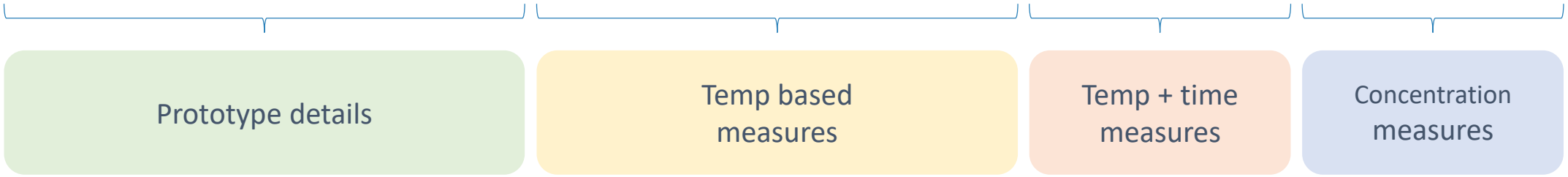
Which month do you celebrate your birthday?

- ○ January
- ○ February
- ○ March
- ○ April
- ○ May
- ○ June
- ○ July

Back to my multi column challenge

Prototype overview example

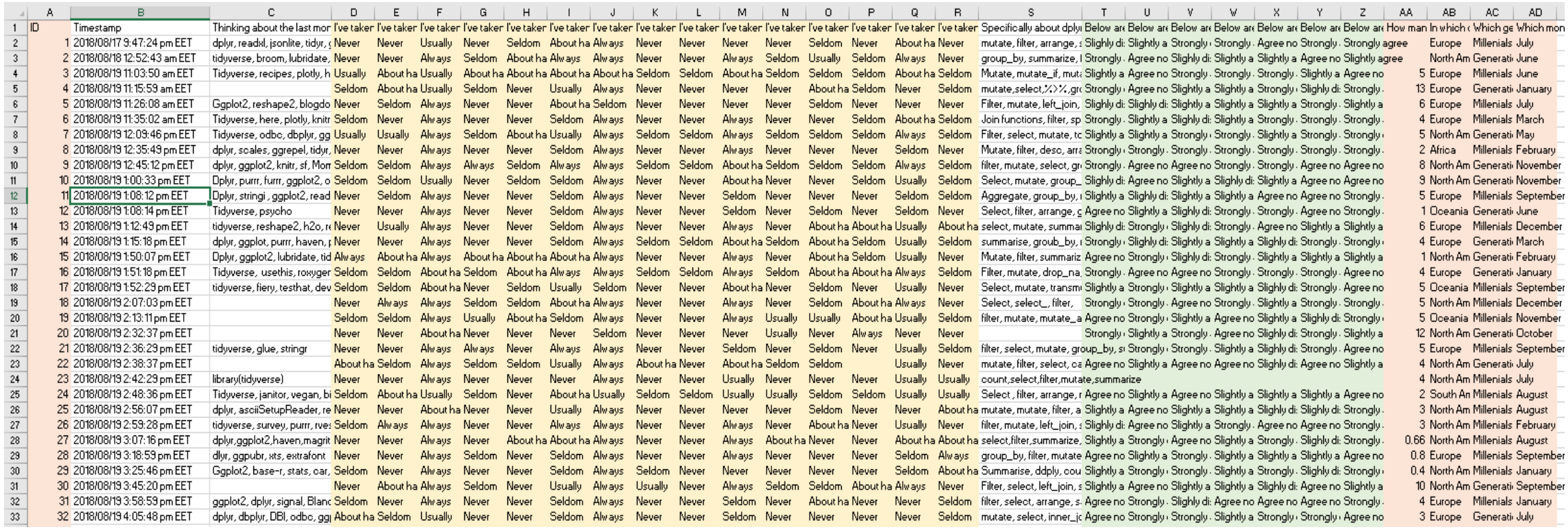
Prototype	RM1	RM2	RM3	...	RM15	Info1	...	Info5	M1_20C	M1_40C	M1_60C	M2_0C	M2_10C	M2_20C	M3_20C_1s	M3_20C_10s	M3_20C_50s	M4_10g/l	M4_20g/l	M4_30g/l
QWE	1	1	0.5	...	1	A	...	1	1	1	1	1	1	1	1	1	1	1	1	1
RTY	2	1	0.6	...	1	B	...	2	2	2	2	2	2	2	2	2	2	2	2	2
UIO	3	1	0.7	...	1	C	...	3	3	3	3	3	3	3	3	3	3	3	3	3
ASD	4	1	0.8	...	1	A	...	4	4	4	4	4	4	4	4	4	4	4	4	4
FGH	5	1	0.9	...	1	B	...	5	5	5	5	5	5	5	5	5	5	5	5	5
HJK	1	1	0.10	...	1	C	...	6	6	6	6	6	6	6	6	6	6	6	6	6
JKL	2	1	0.11	...	1	A	...	7	7	7	7	7	7	7	7	7	7	7	7	7
ZXC	3	1	0.12	...	1	B	...	8	8	8	8	8	8	8	8	8	8	8	8	8
XCV	4	1	0.13	...	1	C	...	9	9	9	9	9	9	9	9	9	9	9	9	9
CVB	5	1	0.5	...	1	A	...	10	10	10	10	10	10	10	10	10	10	10	10	10
VBN	1	1	0.6	...	1	B	...	1	1	1	1	1	1	1	1	1	1	1	1	1
SRT	2	1	0.7	...	1	C	...	2	2	2	2	2	2	2	2	2	2	2	2	2
NEW	3	1	0.8	...	1	A	...	3	3	3	3	3	3	3	3	3	3	3	3	3
LUT	4	1	0.9	...	1	B	...	4	4	4	4	4	4	4	4	4	4	4	4	4
SYL	5	1	0.10	...	1	C	...	5	5	5	5	5	5	5	5	5	5	5	5	5
NEM	1	1	0.11	...	1	A	...	6	6	6	6	6	6	6	6	6	6	6	6	6
AWE	2	1	0.12	...	1	B	...	7	7	7	7	7	7	7	7	7	7	7	7	7
ASD	3	1	0.13	...	1	C	...	8	8	8	8	8	8	8	8	8	8	8	8	8
ERH	4	1	0.5	...	1	A	...	9	9	9	9	9	9	9	9	9	9	9	9	9
XRT	5	1	0.6	...	1	B	...	1	1	1	1	1	1	1	1	1	1	1	1	1
PSJ	1	1	0.7	...	1	C	...	2	2	2	2	2	2	2	2	2	2	2	2	2
ZJR	2	1	0.8	...	1	A	...	3	3	3	3	3	3	3	3	3	3	3	3	3
WKD	3	1	0.9	...	1	B	...	4	4	4	4	4	4	4	4	4	4	4	4	4
WMD	4	1	0.10	...	1	C	...	5	5	5	5	5	5	5	5	5	5	5	5	5



Prototype overview example

Prototype	RM1	RM2	…	RM15	Info1	…	Info5	Temperat	Time	Concentra	M1	M2	M3	M4
QWE	1	1	…	1	A	…	1	20	NA	NA	1	1	NA	NA
QWE	1	1	…	1	A	…	1	40	NA	NA	1	NA	NA	NA
QWE	1	1	…	1	A	…	1	60	NA	NA	1	NA	NA	NA
QWE	1	1	…	1	A	…	1	0	NA	NA	NA	1	NA	NA
QWE	1	1	…	1	A	…	1	10	NA	NA	NA	1	NA	NA
QWE	1	1	…	1	A	…	1	20	1	NA	NA	NA	1	NA
QWE	1	1	…	1	A	…	1	20	10	NA	NA	NA	1	NA
QWE	1	1	…	1	A	…	1	20	50	NA	NA	NA	1	NA
QWE	1	1	…	1	A	…	1	NA	NA	10	NA	NA	NA	1
QWE	1	1	…	1	A	…	1	NA	NA	20	NA	NA	NA	1
QWE	1	1	…	1	A	…	1	NA	NA	30	NA	NA	NA	1
RTY	2	2		1	B	…	2	20	NA	NA	2	2	NA	NA
RTY	2	2	…	1	B	…	2	40	NA					NA
RTY	2	2	…	1	B	…	2	60	NA					NA
RTY	2	2	…	1	B	…	2	0	NA					NA
RTY	2	2	…	1	B	…	2	10	NA					NA
RTY	2	2	…	1	B	…	2	20	1					NA
RTY	2	2	…	1	B	…	2	20	10					NA
RTY	2	2	…	1	B	…	2	20	50					NA
RTY	2	2	…	1	B	…	2	NA	NA					2
RTY	2	2	…	1	B	…	2	NA	NA					2
RTY	2	2	…	1	B	…	2	NA	NA					2

Survey example



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	ID	Timestamp	Thinking about the last mor	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	Specifically about dplyr	Below are	Below are	Below are	Below are	Below are	Below are	Below are	How man	In which c	Which ge	Which mon
2	1	2018/08/17 9:47:24 pm EET	dplyr, readxl, jsonlite, tidyr,	Never	Never	Usually	Never	Seldom	About ha	Always	Never	Never	Never	Never	Seldom	Never	About ha	Never	mutate, filter, arrange,	Slighly di	Slightly a	Strongly	Strongly	Agree no	Strongly	Strongly	agree	Europe	Millenials	July
3	2	2018/08/18 12:52:43 am EET	tidyverse, broom, lubridate,	Never	Never	Always	Seldom	About ha	Always	Always	Never	Never	Always	Seldom	Usually	Seldom	Always	Never	group_by, summarize,	Strongly	Agree no	Slighly di	Slightly a	Slightly a	Agree no	Slightly	agree	North Am	Generati	June
4	3	2018/08/19 11:03:50 am EET	Tidyverse, recipes, plotly, h	Usually	About ha	Usually	About ha	About ha	About ha	About ha	Seldom	Seldom	About ha	Seldom	Seldom	Seldom	About ha	Seldom	Mutate, mutate_if, mut	Slightly a	Agree no	Strongly	Strongly	Strongly	Slightly a	Agree no	5	Europe	Millenials	June
5	4	2018/08/19 11:15:59 am EET		Seldom	About ha	Usually	Seldom	Never	Usually	Always	Never	Never	Never	Never	About ha	Seldom	Never	Seldom	mutate,select,%>%,gr	Strongly	Agree no	Slightly a	Slightly a	Strongly	Slighly di	Strongly	13	Europe	Generati	January
6	5	2018/08/19 11:26:08 am EET	Ggplot2, reshape2, blogdo	Never	Seldom	Always	Never	Never	About ha	Seldom	Never	Never	Never	Never	Seldom	Never	Never	Never	Filter, mutate, left_join,	Slighly di	Slighly di	Slightly a	Slightly a	Slightly a	Strongly	Slightly a	6	Europe	Millenials	July
7	6	2018/08/19 11:35:02 am EET	Tidyverse, here, plotly, knitr	Seldom	Never	Always	Never	Never	Seldom	Always	Never	Never	Always	Never	Never	Seldom	About ha	Seldom	Join functions, filter, sp	Strongly	Slightly a	Slighly di	Slightly a	Strongly	Strongly	Strongly	4	Europe	Millenials	March
8	7	2018/08/19 12:09:46 pm EET	Tidyverse, odbc, dbplyr, gg	Usually	Usually	Always	Seldom	About ha	Usually	Always	Seldom	Seldom	Always	Seldom	Seldom	Seldom	Always	Seldom	Filter, select, mutate, tc	Slightly a	Slightly a	Strongly	Strongly	Slightly a	Strongly	Strongly	5	North Am	Generati	May
9	8	2018/08/19 12:35:49 pm EET	dplyr, scales, ggrepel, tidyr,	Never	Never	Always	Never	Never	Seldom	Always	Never	Never	Always	Never	Never	Never	Seldom	Never	Mutate, filter, desc, arra	Strongly	Strongly	Strongly	Strongly	Strongly	Strongly	Strongly	2	Africa	Millenials	February
10	9	2018/08/19 12:45:12 pm EET	dplyr, ggplot2, knitr, sf, Mon	Seldom	Seldom	Always	Always	Seldom	Always	Always	Seldom	Seldom	About ha	Seldom	Seldom	Seldom	Always	Seldom	filter, mutate, select, gr	Strongly	Agree no	Agree no	Strongly	Strongly	Agree no	Agree no	8	North Am	Generati	November
11	10	2018/08/19 1:00:33 pm EET	Dplyr, purrr, furrr, ggplot2, o	Seldom	Seldom	Usually	Never	Seldom	Seldom	Always	Never	Never	About ha	Never	Never	Seldom	Usually	Seldom	Select, mutate, group_	Slighly di	Agree no	Slightly a	Slighly di	Slightly a	Agree no	Agree no	9	North Am	Generati	November
12	11	2018/08/19 1:08:12 pm EET	Dplyr, stringi , ggplot2, read	Never	Seldom	Always	Never	Never	Seldom	Always	Never	Never	Seldom	Never	Never	Never	Seldom	Seldom	Aggregate, group_by,	Slightly a	Slighly di	Slightly a	Slightly a	Agree no	Agree no	Strongly	5	Europe	Millenials	September
13	12	2018/08/19 1:08:14 pm EET	Tidyverse, psycho	Never	Never	Always	Never	Never	Seldom	Always	Never	Never	Seldom	Never	Seldom	Never	Seldom	Never	Select, filter, arrange, g	Agree no	Slightly a	Slighly di	Strongly	Strongly	Agree no	Strongly	1	Oceania	Generati	June
14	13	2018/08/19 1:12:49 pm EET	tidyverse, reshape2, h2o, re	Never	Usually	Always	Never	Never	Seldom	Always	Never	Never	Always	Never	About ha	About ha	Usually	About ha	select, mutate, summar	Slighly di	Strongly	Slighly di	Strongly	Agree no	Slightly a	Slightly a	6	Europe	Millenials	December
15	14	2018/08/19 1:15:18 pm EET	dplyr, ggplot, purrr, haven,	Never	Never	Always	Never	Never	Seldom	Always	Seldom	Seldom	About ha	Seldom	About ha	Seldom	Usually	Seldom	summarise, groub_by,	Strongly	Slighly di	Slightly a	Slightly a	Slighly di	Strongly	Strongly	4	Europe	Generati	March
16	15	2018/08/19 1:50:07 pm EET	Dplyr, ggplot2, lubridate, tid	Always	About ha	Always	About ha	About ha	About ha	Always	Never	Never	Always	Never	About ha	Seldom	Usually	Never	Mutate, filter, summariz	Agree no	Strongly	Slighly di	Strongly	Slightly a	Slightly a	Slightly a	1	North Am	Generati	February
17	16	2018/08/19 1:51:18 pm EET	Tidyverse, usethis, roxyger	Seldom	Seldom	About ha	Seldom	About ha	Always	Always	Seldom	Seldom	Always	Seldom	About ha	About ha	Always	Seldom	Filter, mutate, drop_na,	Strongly	Agree no	Agree no	Strongly	Strongly	Strongly	Agree no	4	Europe	Generati	January
18	17	2018/08/19 1:52:29 pm EET	tidyverse, fiery, testhat, dev	Seldom	Seldom	About ha	Never	Seldom	Usually	Seldom	Never	Never	About ha	Never	Seldom	Never	Usually	Never	Select, mutate, transm	Slightly a	Strongly	Slightly a	Slightly a	Slighly di	Strongly	Agree no	5	Oceania	Millenials	September
19	18	2018/08/19 2:07:03 pm EET		Never	Always	Always	Seldom	Seldom	About ha	Always	Never	Never	Always	Never	Seldom	About ha	Always	Never	Select, select_, filter,	Strongly	Strongly	Agree no	Strongly	Slightly a	Strongly	Strongly	5	North Am	Millenials	December
20	19	2018/08/19 2:13:11 pm EET		Seldom	Seldom	Always	Usually	About ha	Seldom	Always	Never	Never	Always	Usually	Usually	About ha	Usually	Seldom	filter, mutate, mutate_a	Agree no	Strongly	Slightly a	Agree no	Slightly a	Slighly di	Strongly	5	Oceania	Millenials	November
21	20	2018/08/19 2:32:37 pm EET		Never	Never	About ha	Never	Never	Never	Seldom	Never	Never	Never	Usually	Never	Always	Never	Never		Strongly	Slightly a	Strongly	Agree no	Slighly di	Strongly	Slightly a	12	North Am	Generati	October
22	21	2018/08/19 2:36:29 pm EET	tidyverse, glue, stringr	Never	Never	Always	Always	Never	Always	Always	Never	Never	Seldom	Never	Seldom	Never	Usually	Seldom	filter, select, mutate, group_by, s	Strongly	Strongly	Slightly a	Slighly di	Strongly	Agree no		5	Europe	Millenials	September
23	22	2018/08/19 2:38:37 pm EET		About ha	Seldom	Always	Seldom	Seldom	Usually	Always	About ha	Never	About ha	Seldom	Seldom		Usually	Never	mutate, filter, select, ca	Agree no	Slightly a	Agree no	Agree no	Slighly di	Agree no	Slightly a	4	North Am	Generati	July
24	23	2018/08/19 2:42:29 pm EET	library(tidyverse)	Never	Never	Always	Never	Never	Never	Always	Never	Never	Usually	Never	Never	Never	Usually	Usually	count,select,filter,mutate,summarize								4	North Am	Millenials	July
25	24	2018/08/19 2:48:36 pm EET	Tidyverse, janitor, vegan, bi	Seldom	About ha	Usually	Seldom	Never	About ha	Usually	Seldom	Seldom	Usually	Usually	Seldom	Seldom	Usually	Usually	Select , filter, arrange, r	Agree no	Slightly a	Agree no	Slightly a	Slightly a	Slightly a	Agree no	2	South An	Millenials	August
26	25	2018/08/19 2:56:07 pm EET	dplyr, asciiSetupReader, re	Never	Never	About ha	Never	Never	Usually	Always	Never	Never	Never	Never	Seldom	Never	Never	About ha	mutate, mutate, filter, a	Slightly a	Agree no	Slightly a	Slightly a	Slighly di	Slighly di	Strongly	3	North Am	Millenials	August
27	26	2018/08/19 2:59:28 pm EET	tidyverse, survey, purrr, rves	Seldom	Always	Always	Never	Never	Always	Always	Never	Never	Always	Never	About ha	Never	Usually	Never	filter, mutate, left_join, s	Slighly di	Agree no	Slightly a	Strongly	Agree no	Agree no	Strongly	3	North Am	Millenials	February
28	27	2018/08/19 3:07:16 pm EET	dplyr,ggplot2,haven,magrit	Never	Never	Always	Never	About ha	About ha	Always	Never	Never	Always	About ha	Never	Never	About ha	About ha	select,filter,summarize,	Slightly a	Strongly	Agree no	Slightly a	Strongly	Slighly di	Strongly	0.66	North Am	Millenials	August
29	28	2018/08/19 3:18:59 pm EET	dlyr, ggpubr, xts, extrafont	Never	Never	Always	Seldom	Never	Seldom	Always	Never	Never	Always	Never	Never	Never	Seldom	Always	group_by, filter, mutate	Agree no	Slightly a	Strongly	Slightly a	Slightly a	Slightly a	Agree no	0.8	Europe	Millenials	September
30	29	2018/08/19 3:25:46 pm EET	Ggplot2, base-r, stats, car,	Seldom	Never	Always	Never	Never	Seldom	Always	Seldom	Never	Never	Never	Never	Never	Seldom	About ha	Summarise, ddply, cou	Slightly a	Strongly	Strongly	Slightly a	Strongly	Slighly di	Strongly	0.4	North Am	Millenials	January
31	30	2018/08/19 3:45:20 pm EET		Never	About ha	Always	Seldom	Never	Usually	Always	Usually	Never	Always	Seldom	Seldom	About ha	Always	Never	Filter, select, left_join, s	Slightly a	Agree no	Slightly a	Slightly a	Slighly di	Agree no	Slightly a	10	North Am	Generati	September
32	31	2018/08/19 3:58:59 pm EET	ggplot2, dplyr, signal, Blanc	Seldom	Never	Always	Never	Never	Seldom	Always	Never	Never	Seldom	Never	About ha	Never	Never	Seldom	filter, select, arrange, s	Agree no	Strongly	Slighly di	Agree no	Agree no	Agree no	Strongly	4	Europe	Millenials	January
33	32	2018/08/19 4:05:48 pm EET	dplyr, dbplyr, DBI, odbc, ggp	About ha	Seldom	Usually	Never	Never	Seldom	Always	Never	Never	Seldom	Never	Never	Never	Never	Seldom	mutate, select, inner_jc	Agree no	Strongly	Strongly	Slightly a	Strongly	Strongly	Agree no	3	Europe	Generati	July

Frequency questions

Agreement questions

Demo-graphics

Survey example

1. Clean up all variable names

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	ID	Timestamp	Thinking about the last mor	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	I've taker	Specifically about dplyr	Below are	Below are	Below are	Below are	Below are	Below are	Below are	How man	In which c	Which ge	Which mon
2	1	2018/08/17 9:47:24 pm EET	dplyr, readxl, jsonlite, tidyr,	Never	Never	Usually	Never	Seldom	About ha	Always	Never	Never	Never	Never	Seldom	Never	About ha	Never	mutate, filter, arrange,	Slighly di	Slightly a	Strongly	Strongly	Agree no	Strongly	Strongly	agree	Europe	Millenials	July
3	2	2018/08/18 12:52:43 am EET	tidyverse, broom, lubridate,	Never	Never	Always	Seldom	About ha	Always	Always	Never	Never	Always	Seldom	Usually	Seldom	Always	Never	group_by, summarize,	Strongly	Agree no	Slighly di	Slightly a	Slightly a	Agree no	Slightly	agree	North Am	Generati	June
4	3	2018/08/19 11:03:50 am EET	Tidyverse, recipes, plotly, h	Usually	About ha	Usually	About ha	About ha	About ha	About ha	Seldom	Seldom	About ha	Seldom	Seldom	Seldom	About ha	Seldom	Mutate, mutate_if, mut	Slightly a	Agree no	Strongly	Strongly	Strongly	Slightly a	Agree no	5	Europe	Millenials	June
5	4	2018/08/19 11:15:59 am EET		Seldom	About ha	Usually	Seldom	Never	Usually	Always	Never	Never	Never	Never	About ha	Seldom	Never	Seldom	mutate,select,%>%,gr	Strongly	Agree no	Slightly a	Slightly a	Strongly	Slighly di	Strongly	13	Europe	Generati	January
6	5	2018/08/19 11:26:08 am EET	Ggplot2, reshape2, blogdo	Never	Seldom	Always	Never	Never	About ha	Seldom	Never	Never	Never	Never	Seldom	Never	Never	Never	Filter, mutate, left_join,	Slighly di	Slighly di	Slightly a	Slightly a	Slightly a	Strongly	Slightly a	6	Europe	Millenials	July
7	6	2018/08/19 11:35:02 am EET	Tidyverse, here, plotly, knitr	Seldom	Never	Always	Never	Never	Seldom	Always	Never	Never	Always	Never	Never	Seldom	About ha	Seldom	Join functions, filter, sp	Strongly	Slightly a	Slighly di	Slightly a	Strongly	Strongly	Strongly	4	Europe	Millenials	March
8	7	2018/08/19 12:09:46 pm EET	Tidyverse, odbc, dbplyr, gg	Usually	Usually	Always	Seldom	About ha	Usually	Always	Seldom	Seldom	Always	Seldom	Seldom	Seldom	Always	Seldom	Filter, select, mutate, tc	Slightly a	Slightly a	Strongly	Strongly	Slightly a	Strongly	Strongly	5	North Am	Generati	May
9	8	2018/08/19 12:35:49 pm EET	dplyr, scales, ggrepel, tidyr,	Never	Never	Always	Never	Never	Seldom	Always	Never	Never	Always	Never	Never	Never	Seldom	Never	Mutate, filter, desc, arra	Strongly	Strongly	Strongly	Strongly	Strongly	Strongly	Strongly	2	Africa	Millenials	February
10	9	2018/08/19 12:45:12 pm EET	dplyr, ggplot2, knitr, sf, Mon	Seldom	Seldom	Always	Always	Seldom	Always	Always	Seldom	Seldom	About ha	Seldom	Seldom	Seldom	Always	Seldom	filter, mutate, select, gr	Strongly	Agree no	Agree no	Strongly	Strongly	Agree no	Agree no	8	North Am	Generati	November
11	10	2018/08/19 1:00:33 pm EET	Dplyr, purrr, furrr, ggplot2, o	Seldom	Seldom	Usually	Never	Seldom	Seldom	Always	Never	Never	About ha	Never	Never	Seldom	Usually	Seldom	Select, mutate, group_	Slighly di	Agree no	Slightly a	Slighly di	Slightly a	Agree no	Agree no	9	North Am	Generati	November
12	11	2018/08/19 1:08:12 pm EET	Dplyr, stringi , ggplot2, read	Never	Seldom	Always	Never	Never	Seldom	Always	Never	Never	Seldom	Never	Never	Never	Seldom	Seldom	Aggregate, group_by,	Slightly a	Slighly di	Slightly a	Slightly a	Agree no	Agree no	Strongly	5	Europe	Millenials	September
13	12	2018/08/19 1:08:14 pm EET	Tidyverse, psycho	Never	Never	Always	Never	Never	Seldom	Always	Never	Never	Seldom	Never	Seldom	Never	Seldom	Never	Select, filter, arrange, g	Agree no	Slightly a	Slighly di	Strongly	Strongly	Agree no	Strongly	1	Oceania	Generati	June
14	13	2018/08/19 1:12:49 pm EET	tidyverse, reshape2, h2o, re	Never	Usually	Always	Never	Never	Seldom	Always	Never	Never	Always	Never	About ha	About ha	Usually	About ha	select, mutate, summar	Slighly di	Strongly	Slighly di	Strongly	Agree no	Slightly a	Slightly a	6	Europe	Millenials	December
15	14	2018/08/19 1:15:18 pm EET	dplyr, ggplot, purrr, haven, p	Never	Never	Always	Never	Never	Seldom	Always	Seldom	Seldom	About ha	Seldom	About ha	Seldom	Usually	Seldom	summarise, groub_by,	Strongly	Slighly di	Slightly a	Slightly a	Slighly di	Strongly	Strongly	4	Europe	Generati	March
16	15	2018/08/19 1:50:07 pm EET	Dplyr, ggplot2, lubridate, tid	Always	About ha	Always	About ha	About ha	About ha	Always	Never	Never	Always	Never	About ha	Seldom	Usually	Never	Mutate, filter, summariz	Agree no	Strongly	Slighly di	Strongly	Slightly a	Slightly a	Slightly a	1	North Am	Generati	February
17	16	2018/08/19 1:51:18 pm EET	Tidyverse, usethis, roxyger	Seldom	Seldom	About ha	Seldom	About ha	Always	Always	Seldom	Seldom	Always	Seldom	About ha	About ha	Always	Seldom	Filter, mutate, drop_na,	Strongly	Agree no	Agree no	Strongly	Strongly	Strongly	Agree no	4	Europe	Generati	January
18	17	2018/08/19 1:52:29 pm EET	tidyverse, fiery, testhat, dev	Seldom	Seldom	About ha	Never	Seldom	Usually	Seldom	Never	Never	About ha	Never	Seldom	Never	Usually	Never	Select, mutate, transm	Slightly a	Strongly	Slightly a	Slightly a	Slighly di	Strongly	Agree no	5	Oceania	Millenials	September
19	18	2018/08/19 2:07:03 pm EET		Never	Always	Always	Seldom	Seldom	About ha	Always	Never	Never	Always	Never	Seldom	About ha	Always	Never	Select, select_, filter,	Strongly	Strongly	Agree no	Strongly	Slightly a	Strongly	Strongly	5	North Am	Millenials	December
20	19	2018/08/19 2:13:11 pm EET		Seldom	Seldom	Always	Usually	About ha	Seldom	Always	Never	Never	Always	Usually	Usually	About ha	Usually	Seldom	filter, mutate, mutate_a	Agree no	Strongly	Slightly a	Agree no	Slightly a	Slighly di	Strongly	5	Oceania	Millenials	November
21	20	2018/08/19 2:32:37 pm EET		Never	Never	About ha	Never	Never	Never	Seldom	Never	Never	Never	Usually	Never	Always	Never	Never		Strongly	Slightly a	Strongly	Agree no	Slighly di	Strongly	Slightly a	12	North Am	Generati	October
22	21	2018/08/19 2:36:29 pm EET	tidyverse, glue, stringr	Never	Never	Always	Always	Never	Always	Always	Never							Seldom	filter, select, mutate, group_by, s	Strongly	Strongly	Slightly a	Slighly di	Strongly	Agree no	5	Europe	Millenials	September
23	22	2018/08/19 2:38:37 pm EET		About ha	Seldom	Always	Seldom	Seldom	Usually	Always	About							Never	mutate, filter, select, ca	Agree no	Slightly a	Agree no	Agree no	Slighly di	Agree no	Slightly a	4	North Am	Generati	July
24	23	2018/08/19 2:42:29 pm EET	library(tidyverse)	Never	Never	Always	Never	Never	Never	Always	Never							Usually	count,select,filter,mutate,summarize								4	North Am	Millenials	July
25	24	2018/08/19 2:48:36 pm EET	Tidyverse, janitor, vegan, bi	Seldom	About ha	Usually	Seldom	Never	About ha	Usually	Seldo							Usually	Select , filter, arrange, r	Agree no	Slightly a	Agree no	Slightly a	Slightly a	Slightly a	Agree no	2	South An	Millenials	August
26	25	2018/08/19 2:56:07 pm EET	dplyr, asciiSetupReader, re	Never	Never	About ha	Never	Never	Usually	Always	Never							About ha	mutate, mutate, filter, a	Slighly a	Agree no	Slightly a	Slightly a	Slighly di	Slighly di	Strongly	3	North Am	Millenials	August
27	26	2018/08/19 2:59:28 pm EET	tidyverse, survey, purrr, rves	Seldom	Always	Always	Never	Never	Always	Always	Never							Never	filter, mutate, left_join, s	Slighly di	Agree no	Slightly a	Strongly	Agree no	Agree no	Strongly	3	North Am	Millenials	February
28	27	2018/08/19 3:07:16 pm EET	dplyr,ggplot2,haven,magrit	Never	Never	Always	Never	About ha	About ha	Always	Never							About ha	select,filter,summarize,	Slightly a	Strongly	Agree no	Slightly a	Strongly	Slighly di	Strongly	0.66	North Am	Millenials	August
29	28	2018/08/19 3:18:59 pm EET	dlyr, ggpubr, xts, extrafont	Never	Never	Always	Seldom	Never	Seldom	Always	Never							Always	group_by, filter, mutate	Agree no	Slightly a	Strongly	Slightly a	Slightly a	Slightly a	Agree no	0.8	Europe	Millenials	September
30	29	2018/08/19 3:25:46 pm EET	Ggplot2, base-r, stats, car,	Seldom	Never	Always	Never	Never	Seldom	Always	Seldo							About ha	Summarise, ddply, cou	Slightly a	Strongly	Strongly	Slightly a	Strongly	Slighly di	Strongly	0.4	North Am	Millenials	January
31	30	2018/08/19 3:45:20 pm EET		Never	About ha	Always	Seldom	Never	Usually	Always	Usually	Never	Always	Seldom	Seldom	About ha	Always	Never	Filter, select, left_join, s	Slightly a	Agree no	Slightly a	Slightly a	Slighly di	Agree no	Slightly a	10	North Am	Generati	September
32	31	2018/08/19 3:58:59 pm EET	ggplot2, dplyr, signal, Blanc	Seldom	Never	Always	Never	Never	Seldom	Always	Never	Never	Seldom	Never	About ha	Never	Never	Seldom	filter, select, arrange, s	Agree no	Strongly	Slighly di	Agree no	Agree no	Agree no	Strongly	4	Europe	Millenials	January
33	32	2018/08/19 4:05:48 pm EET	dplyr, dbplyr, DBI, odbc, ggi	About ha	Seldom	Usually	Never	Never	Seldom	Always	Never	Never	Seldom	Never	Never	Never	Never	Seldom	mutate, select, inner_jc	Agree no	Strongly	Strongly	Slightly a	Strongly	Strongly	Agree no	3	Europe	Generati	July

2. Recode all observations

3. Summarise all related variables

What are my options?

What are my options?

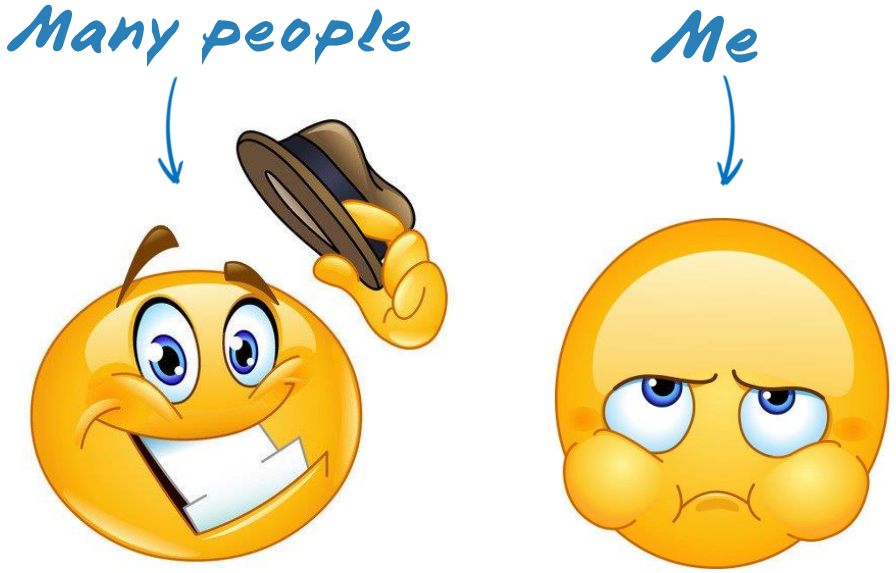
- Manual column by column

What are my options?

- Manual column by column
- Iteration or loops

What are my options?

- Manual column by column
- Iteration or loops
- Grabbing a selection and gather, mutate, group_by and summarise



Package ‘dplyr’

June 29, 2018

Type Package

Title A Grammar of Data Manipulation

Version 0.7.6

Description A fast, consistent tool for working with data frame like objects, both in memory and out of memory.

License MIT + file LICENSE

URL http://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr

BugReports https://github.com/tidyverse/dplyr/issues

Depends R (>= 3.1.2)

Imports assertthat (>= 0.2.0), bindrcpp (>= 0.2.0.9000), glue (>= 1.1.1), magrittr (>= 1.5), methods, pkgconfig (>= 2.0.1), R6 (>= 2.2.2), Rcpp (>= 0.12.15), rlang (>= 0.2.0), tibble (>= 1.3.1), tidyselect (>= 0.2.3), utils

Suggests bit64 (>= 0.9.7), callr, covr (>= 3.0.1), DBI (>= 0.7.14), dbplyr (>= 1.2.0), dtplyr (>= 0.0.2), ggplot2 (>= 2.2.1), hms (>= 0.4.1), knitr (>= 1.19), Lahman (>= 3.0-1), lubridate, MASS, mgcv (>= 1.8.23), microbenchmark (>= 1.4.4), nycflights13 (>= 0.2.2), rmarkdown (>= 1.8), RMySQL (>= 0.10.13), RPostgreSQL (>= 0.6.2), RSQLite (>= 2.0), testthat (>= 2.0.0), withr (>= 2.1.1)

LinkingTo BH (>= 1.58.0-1), bindrcpp (>= 0.2.0.9000), plogr (>= 0.1.10), Rcpp (>= 0.12.15)

VignetteBuilder knitr

Encoding UTF-8

LazyData yes

RoxygenNote 6.0.1.9000

NeedsCompilation yes

Author Hadley Wickham [aut, cre] (<https://orcid.org/0000-0003-4757-117X>), Romain François [aut] (<https://orcid.org/0000-0002-2444-4226>), Lionel Henry [aut], Kirill Müller [aut] (<https://orcid.org/0000-0002-1416-3412>), RStudio [cph, fnd]

What are my options?

- Manual column by column
- Iteration or loops
- Grabbing a piece and making it long
- Scoped dplyr functions

Option 4: Scoped dplyr functions

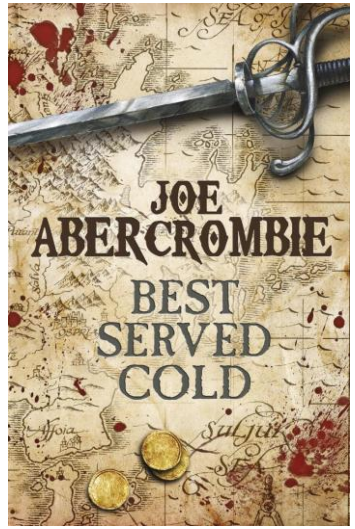
```
clean_data <- rawdata %>%
  select_at(vars(contains("quotes")), ~(str_replace(., "Below .+\\(", "quote_"
  rename_all(~str_remove_all(., "\\W")) %>%
  mutate_all(~if_else(. %in% c("Strongly agree", "Slightly agree"), 1, 0))

summary_data <- clean_data %>%
  summarise_all(mean)
```

Option 4: Scoped dplyr functions

```
summary_data %>%
  gather("quote", "proportion_agree") %>%
  arrange(desc(proportion_agree))
```

```
## # A tibble: 7 x 2
##   quote                  proportion_agree
##   <chr>                             <dbl>
## 1 quote_JoeAbercrombie              0.706
## 2 quote_JimCarrey                   0.611
## 3 quote_MarkTwain                   0.579
## 4 quote_GeorgeRRMartin              0.556
## 5 quote_DouglasAdams                0.548
## 6 quote_Yoda                        0.460
## 7 quote_EvelynWaugh                 0.325
```

"You should laugh every moment you live, for you'll find it decidedly difficult afterwards."



Picture by Tan Danh via pexels.com

Scoped dplyr functions

Scoped dplyr functions

*dplyr*_all

do something to **all variables**

*dplyr*_if

do something to a **set of variables** that were selected based on their **characteristics**

*dplyr*_at

do something to a **set of variables** that were selected based on their variable **names**