

Basic Data Summaries in R Markdown

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

First steps

Install packages you will need, unless you already have them

Load the packages to make the functions available

Load a data set from CSV file

Data Summary

Select the variables you are interested in, removing things like the UUID Then make a summary table using `tbl_summary` command from `gtsummary` package

```
df<-select(df, gender,age,people_in_home,education)
tbl_summary(df,missing = "always")
```

Characteristic	N = 4,541
gender	
female	2,941 (65%)
male	1,515 (33%)
NA	60 (1.3%)
other	25 (0.6%)
Unknown	0
age	43 (35, 51)
Unknown	0
people_in_home	5 (5, 6)
Unknown	0
education	
alevel	313 (6.9%)
do_not_know	39 (0.9%)
further	1,157 (25%)
gcse	498 (11%)
higher	1,171 (26%)
NA	38 (0.8%)
postgrad	1,279 (28%)
primary	46 (1.0%)
Unknown	0

Or split the same data by a variable

In this example I used gender as the basis of the split. Use `fct_explicit_na` command from `forcats` package to make NAs explicit factor of the grouping variable

```
df$gender<-fct_explicit_na(df$gender)
tbl_summary(df,by = "gender",missing = "always")
```

Characteristic	female, N = 2,941	male, N = 1,515	NA, N = 60	other, N = 25
age	43 (35, 51)	44 (35, 51)	41 (30, 48)	42 (38, 48)
Unknown	0	0	0	0
people_in_home	5 (5, 6)	5 (5, 6)	5 (1, 5)	6 (5, 9)
Unknown	0	0	0	0
education				
alevel	196 (6.7%)	113 (7.5%)	3 (5.0%)	1 (4.0%)
do_not_know	23 (0.8%)	15 (1.0%)	1 (1.7%)	0 (0%)
further	761 (26%)	381 (25%)	10 (17%)	5 (20%)
gcse	323 (11%)	164 (11%)	6 (10%)	5 (20%)
higher	749 (25%)	400 (26%)	16 (27%)	6 (24%)
NA	26 (0.9%)	12 (0.8%)	0 (0%)	0 (0%)
postgrad	835 (28%)	414 (27%)	22 (37%)	8 (32%)
primary	28 (1.0%)	16 (1.1%)	2 (3.3%)	0 (0%)
Unknown	0	0	0	0

Draw a chart using ggplot2

```
ggplot(data = df,aes(x=education))+geom_bar()
```

