bootcamp-survey

Rick Gilmore

2017-08-16 12:22:55

Table of Contents

## Goals

* Download and clean data from 2017 R Bootcamp Survey
* Visualize data
* Prepare reports in ioslides\_presentation, pdf\_document, and word\_document formats

## Preliminaries

Load required packages.

library(tidyverse)  
library(googlesheets)

## Load data and examine

The survey data are stored in a [Google Sheet](https://docs.google.com/spreadsheets/d/1Ay56u6g4jyEEdlmV2NHxTLBlcjI2gHavta-Ik0kGrpg/edit#gid=896447063). We'll use the googlesheets package to open it and create a data frame. Documentation about the package can be found [here](https://cran.r-project.org/web/packages/googlesheets/vignettes/basic-usage.html).

There are some idiosyncrasies in using the googlesheets package in an R Markdown document, so I created a separate R script, Get\_bootcamp\_googlesheet.R to extract the survey data and save it to a CSV under data/survey.csv. We can then just load this file.

# Created test data set for testing.  
# survey <- read\_csv("../data/survey.csv")  
survey <- read\_csv("../data/survey-test.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## X1 = col\_integer(),  
## Timestamp = col\_datetime(format = ""),  
## R\_exp = col\_character(),  
## GoT = col\_integer(),  
## Age\_yrs = col\_integer(),  
## Sleep\_hrs = col\_double(),  
## Fav\_date = col\_date(format = ""),  
## Tidy\_data = col\_character()  
## )

survey

## # A tibble: 50 x 8  
## X1 Timestamp R\_exp GoT Age\_yrs Sleep\_hrs Fav\_date  
## <int> <dttm> <chr> <int> <int> <dbl> <date>  
## 1 1 2017-08-15 10:55:01 none 3 52 7.569531 2017-08-15  
## 2 2 2017-08-15 10:55:01 none 3 53 7.742731 2017-08-15  
## 3 3 2017-08-15 10:55:01 some 7 31 6.236837 2017-08-15  
## 4 4 2017-08-15 10:55:01 lots 4 49 8.460097 2017-08-15  
## 5 5 2017-08-15 10:55:01 limited 4 43 7.360005 2017-08-15  
## 6 6 2017-08-15 10:55:01 pro 5 39 8.455450 2017-08-15  
## 7 7 2017-08-15 10:55:01 lots 3 46 8.704837 2017-08-15  
## 8 8 2017-08-15 10:55:01 limited 7 26 9.035104 2017-08-15  
## 9 9 2017-08-15 10:55:01 none 4 44 7.391074 2017-08-15  
## 10 10 2017-08-15 10:55:01 some 4 45 8.504955 2017-08-15  
## # ... with 40 more rows, and 1 more variables: Tidy\_data <chr>

The str() or 'structure' command is great to see what you've got.

str(survey)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 50 obs. of 8 variables:  
## $ X1 : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Timestamp: POSIXct, format: "2017-08-15 10:55:01" "2017-08-15 10:55:01" ...  
## $ R\_exp : chr "none" "none" "some" "lots" ...  
## $ GoT : int 3 3 7 4 4 5 3 7 4 4 ...  
## $ Age\_yrs : int 52 53 31 49 43 39 46 26 44 45 ...  
## $ Sleep\_hrs: num 7.57 7.74 6.24 8.46 7.36 ...  
## $ Fav\_date : Date, format: "2017-08-15" "2017-08-15" ...  
## $ Tidy\_data: chr "Yes" "No" "Yes" "No" ...  
## - attr(\*, "spec")=List of 2  
## ..$ cols :List of 8  
## .. ..$ X1 : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ Timestamp:List of 1  
## .. .. ..$ format: chr ""  
## .. .. ..- attr(\*, "class")= chr "collector\_datetime" "collector"  
## .. ..$ R\_exp : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## .. ..$ GoT : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ Age\_yrs : list()  
## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  
## .. ..$ Sleep\_hrs: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_double" "collector"  
## .. ..$ Fav\_date :List of 1  
## .. .. ..$ format: chr ""  
## .. .. ..- attr(\*, "class")= chr "collector\_date" "collector"  
## .. ..$ Tidy\_data: list()  
## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  
## ..$ default: list()  
## .. ..- attr(\*, "class")= chr "collector\_guess" "collector"  
## ..- attr(\*, "class")= chr "col\_spec"

Notice that the get-bootcamp-googlesheet.R script changed the names of the variables a bit. We may also want to modify the levels of the R\_exp variable to make it an ordered factor.

(survey\_responses <- unique(survey$R\_exp))

## [1] "none" "some" "lots" "limited" "pro"

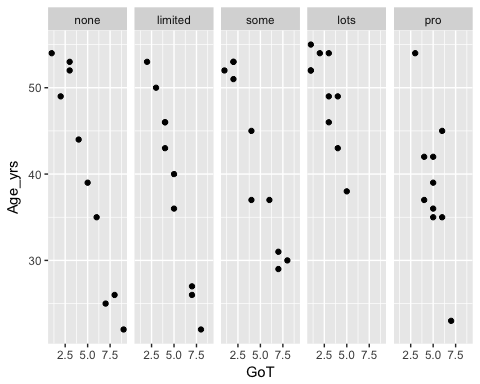
This shows us the different survey response values.

survey$R\_exp <- ordered(survey$R\_exp, levels=c("none",  
 "limited",  
 "some",  
 "lots",  
 "pro"))

## Visualization and analysis

Now, we can ask important questions.

got\_vs\_r\_exp <- survey %>%  
 ggplot() +  
 aes(x=GoT, y=Age\_yrs) +  
 facet\_grid(. ~ R\_exp) +  
 geom\_point()  
got\_vs\_r\_exp



Notice that I sometimes put a label like got-vs-r-exp in the brackets for a given 'chunk' of R code. The main reasons to do this are:

* It sometimes makes it easier to debug your code.
* In some cases, you can have this 'chunk' name serve as the file name for a figure you generate within a chunk.
* In a bit, we'll see how these chunk names are useful for making tables, figures, and equations that generate their own numbers.