bootcamp-survey

Rick Gilmore

2017-08-16 12:22:47

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. We'll use the `googlesheets` package to open it and create a data frame. Documentation about the package can be found here.

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"
```

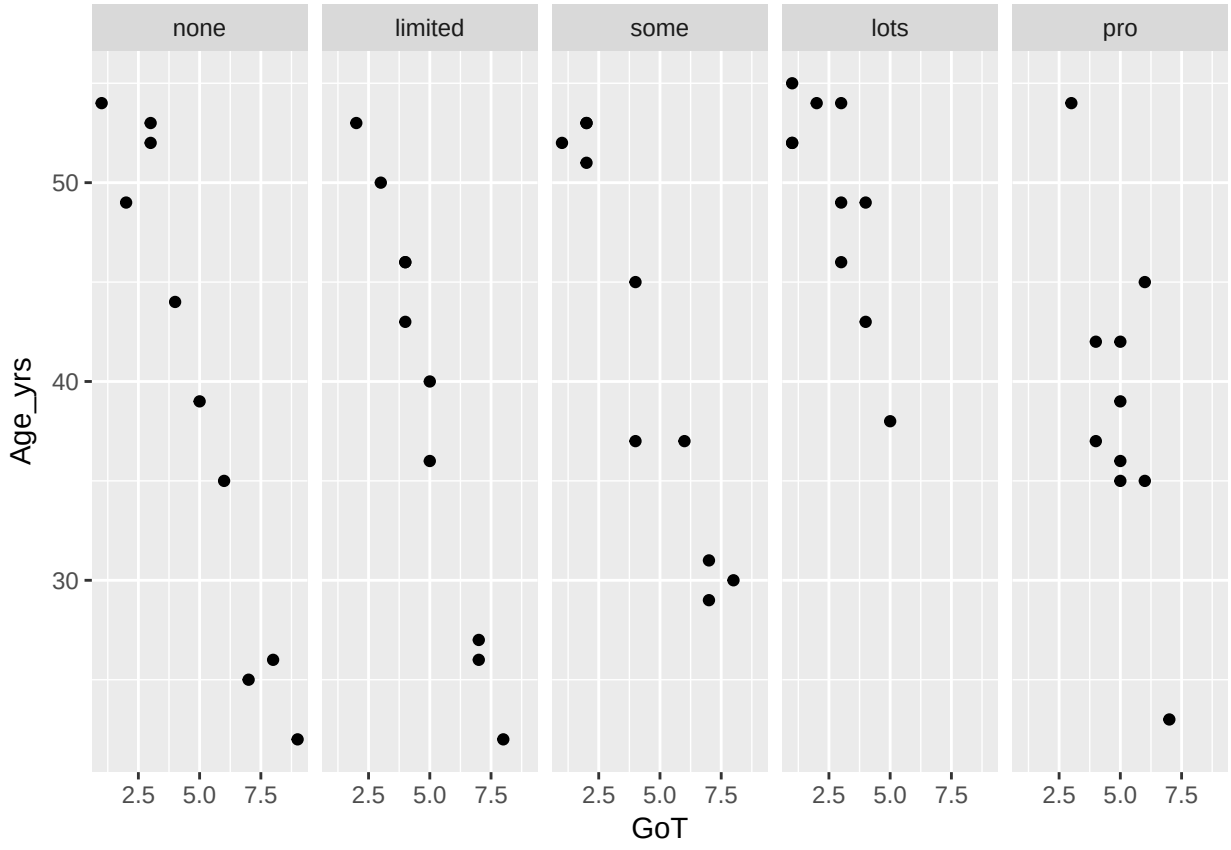
This shows us the different survey response values.

```
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.