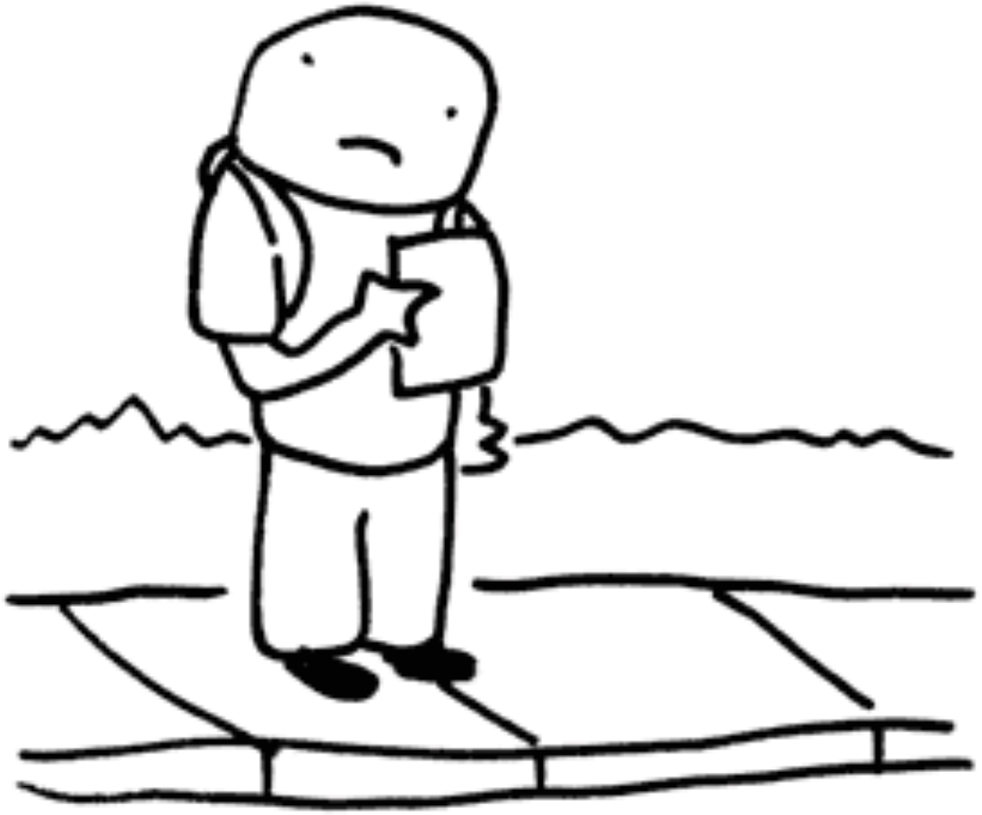
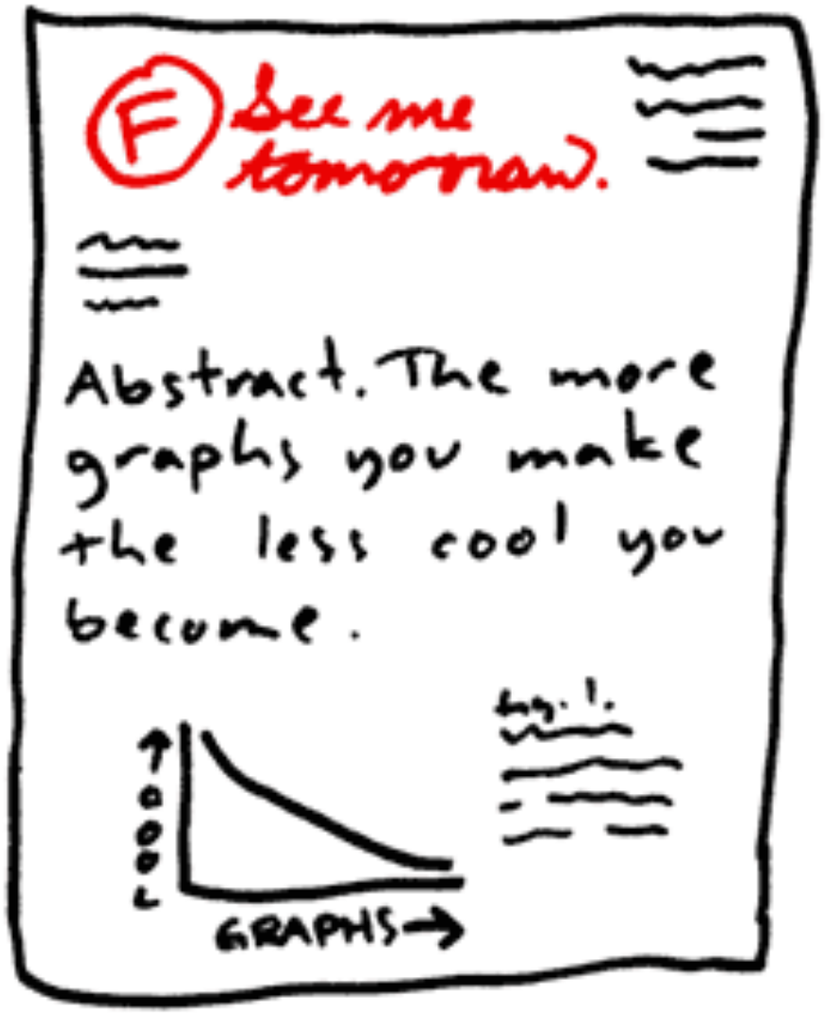
Visualization 2



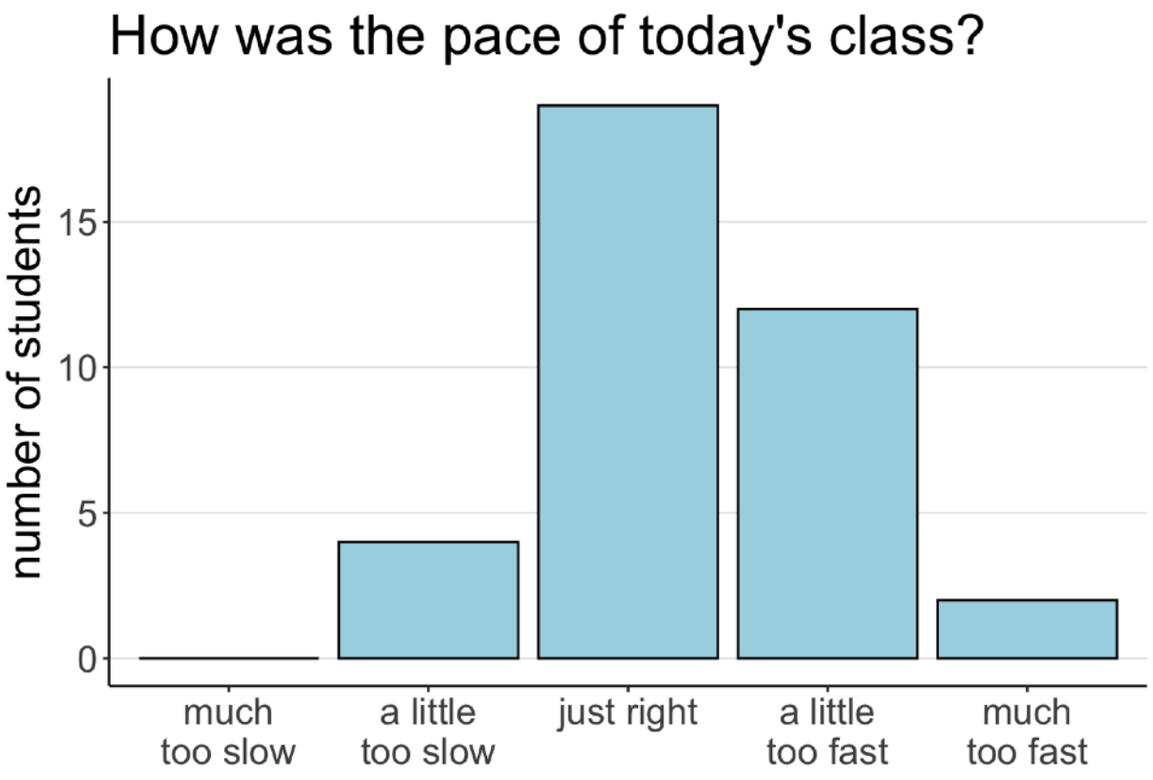


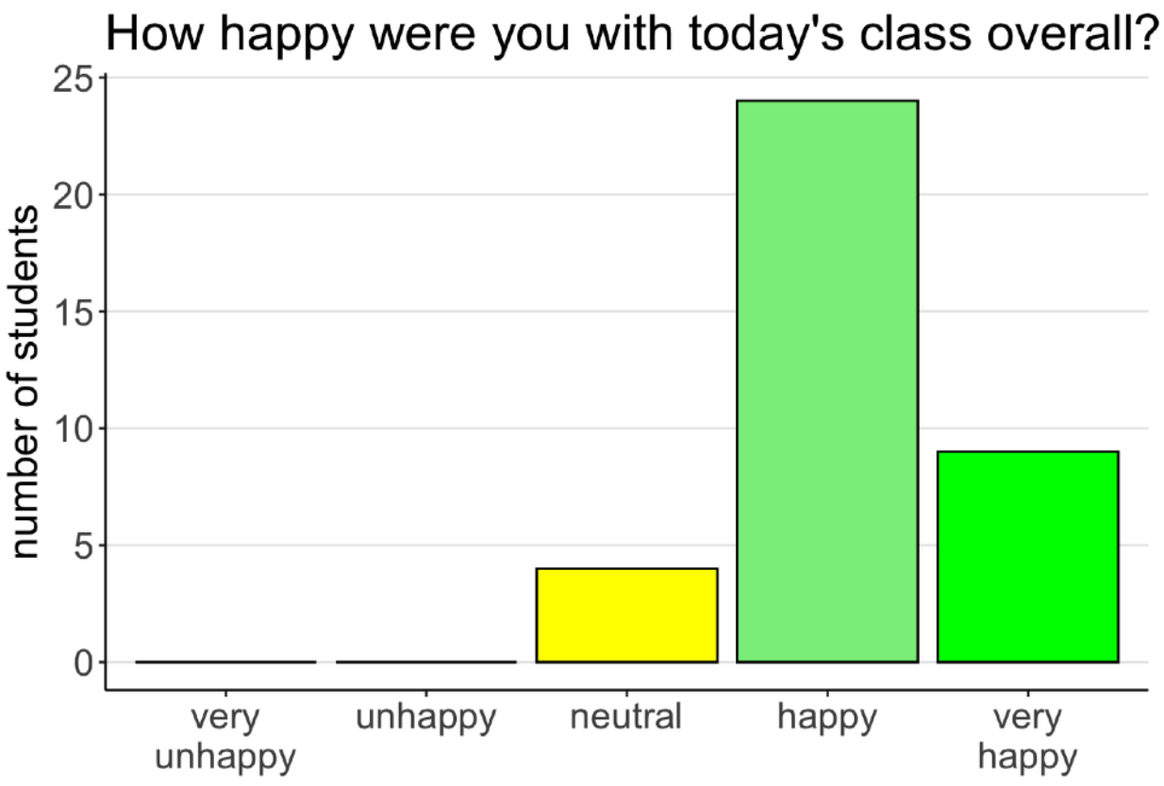
Toothpaste For Dinner.com

01/10/2020

Your feedback

Your feedback





take a look at the datacamp tutorials and other resources mentioned in the RMarkdown scripts

Your feedback

I enjoyed the detailed explanations for the reasons for errors in R. I also thought the practice sets were a good way to consolidate the newly learned concepts. **Please I would like the notes to always be available a day before the class.**

I'll try to make them available as early as possible (might still tweak small things the day of class)

Your feedback

I really liked the parts where we did “hands-on” exercises and wrote code ourselves. **I wish we could have spent a little more time (as a group) going over the “correct” code**, because I felt that the answer part went a little too quickly for those of us who made mistakes.

I'll spend more time on going over the solutions together

Your feedback

It may be nice to actually have more individual/group exercises. I usually find that this kind of material is hard to learn until I have the opportunity to apply it.

come to homework section and application section!

Your feedback

I was surprised by how much overlap there was between the course material and the course readings posted on the website (i.e. word-for-word). **Should I assume moving forward that we don't need to complete the readings before class, but rather that they are most helpful for reviewing anything we missed / need clarity on after a class?**

yes, that's right -- sorry for the confusion!

Friday, January 10th: Visualization II

Content:

- Deciding what plot is appropriate for what kind of data.
- Customizing plots: Take a sad plot and make it better.
- Saving plots.
- Making figure panels.
- Debugging.
- Making animations.
- Defining snippets.

Resources:

- Cheatsheet shiny

Datacamp:

- ggplot part 3
- Shiny 1
- Shiny 2

Reading:

- Course notes: Visualization 2
- Data visualization (#4)
- Data visualization (#8)
- R for Data Science (#27)

Your feedback

Just a heads up that I noticed today before class, two of the data camp courses listed on the course website under the first day (RStudio IDE 1, RStudio IDE 2) are archived on data camp and are no longer available.

thanks! i've removed the broken links from the materials

Your feedback

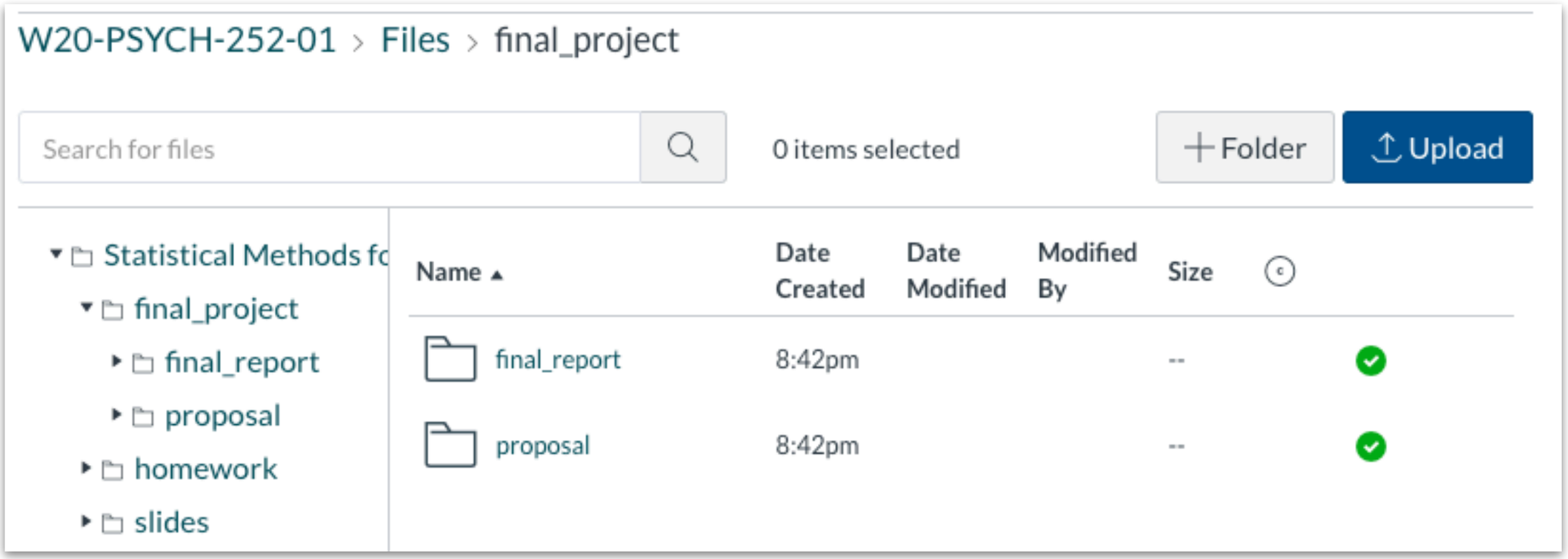
I still didn't figure out how to do an R Project properly. Am I supposed to save the file from Canvas into the R project I created? What if I create a new file? Still not quite sure how to use R Project properly.

the files on Canvas contain an .Rproj file you can open up that file, and then navigate to the RMarkdown file within RStudio's Files browser

the idea is to have one .Rproj that can contain many .r or .rmd files

Final projects

Final projects



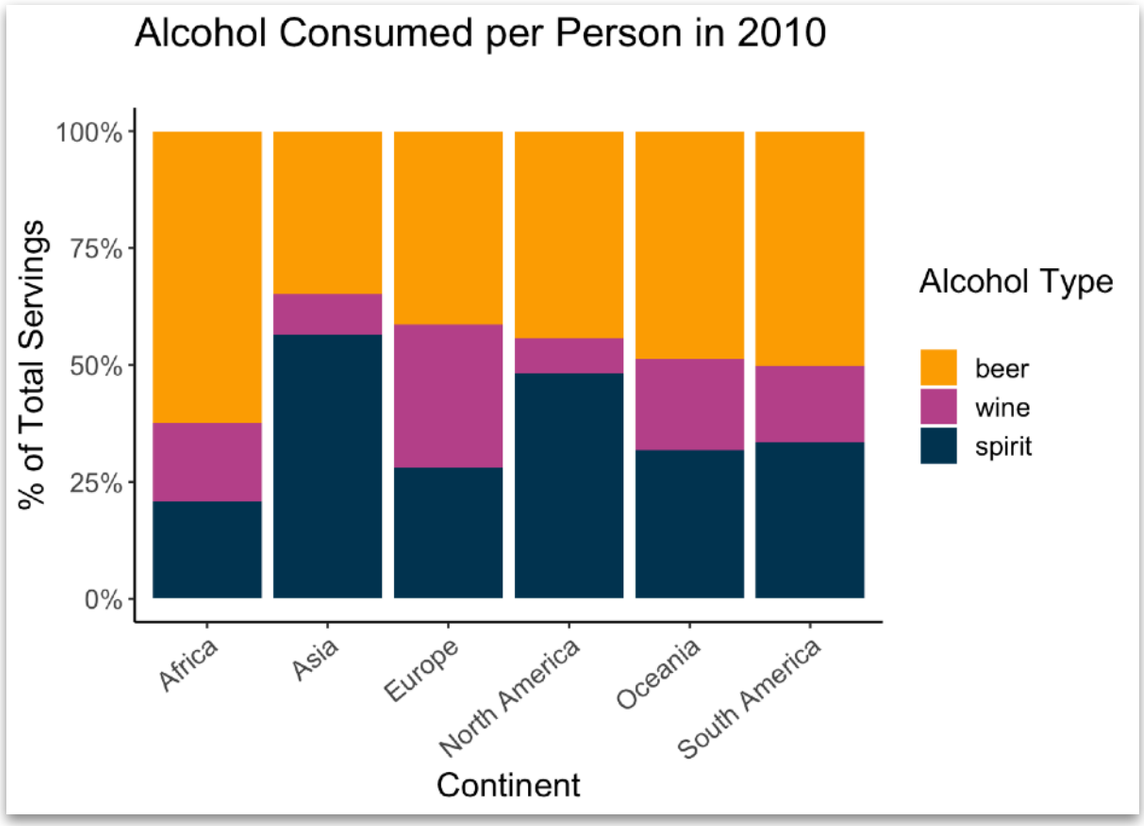
check out examples of proposals and final reports on Canvas

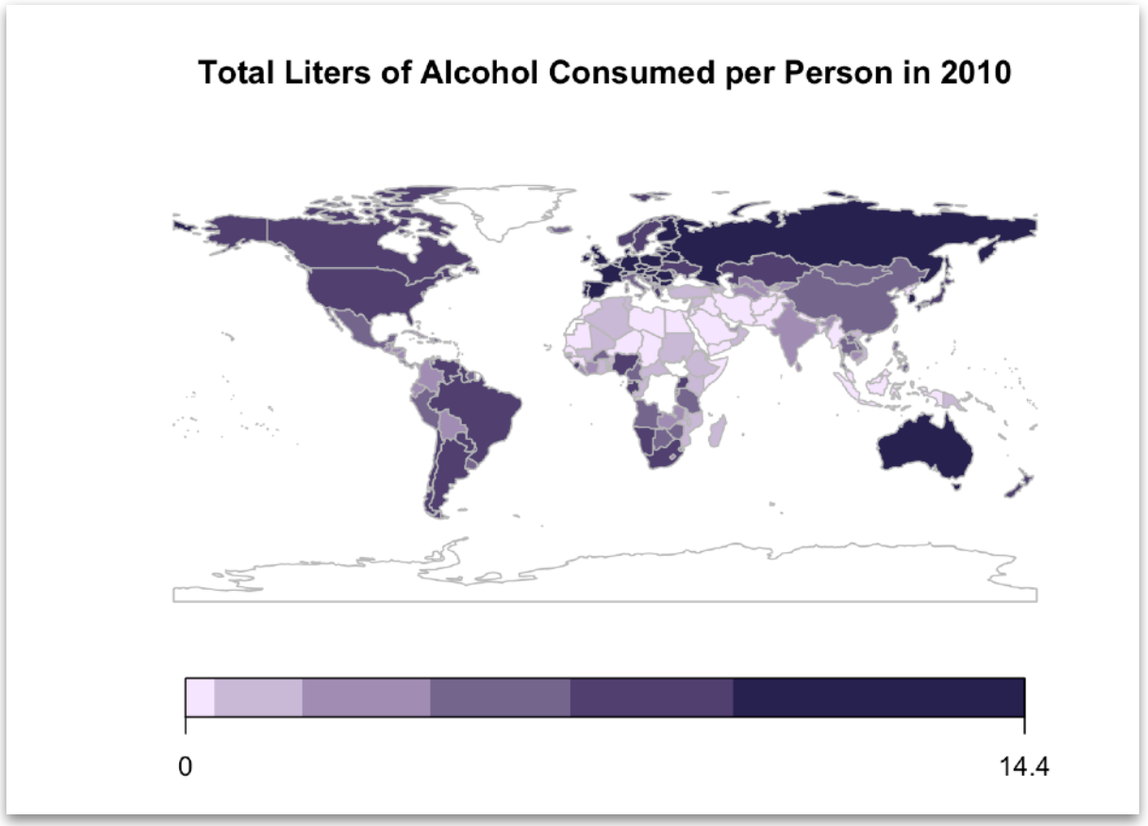
Homework

Homework

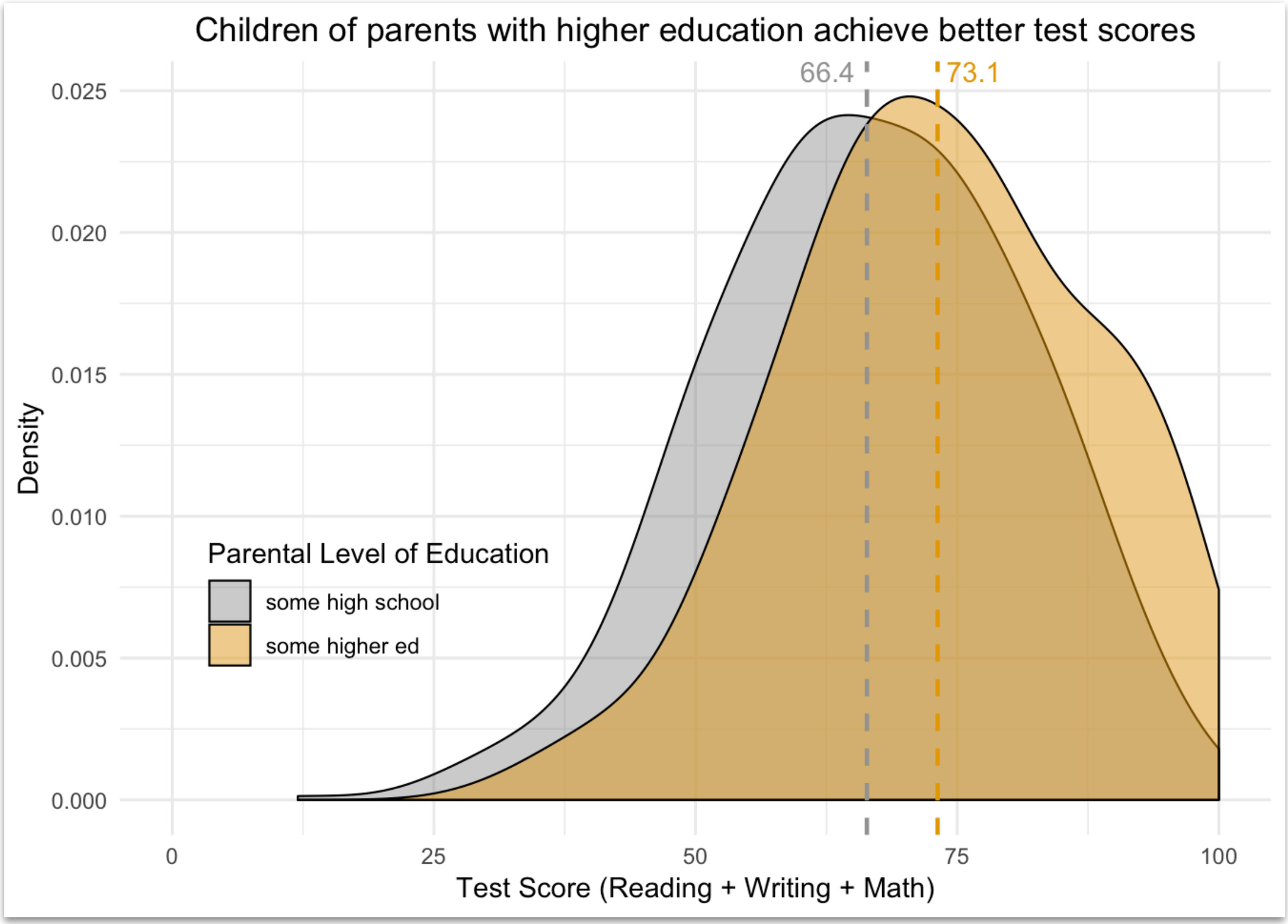
In this homework, **you'll write a short blog post** about a data set. Your goal is to tell us something interesting using a well-crafted, thoughtfully-prepared data graphic.

Homework

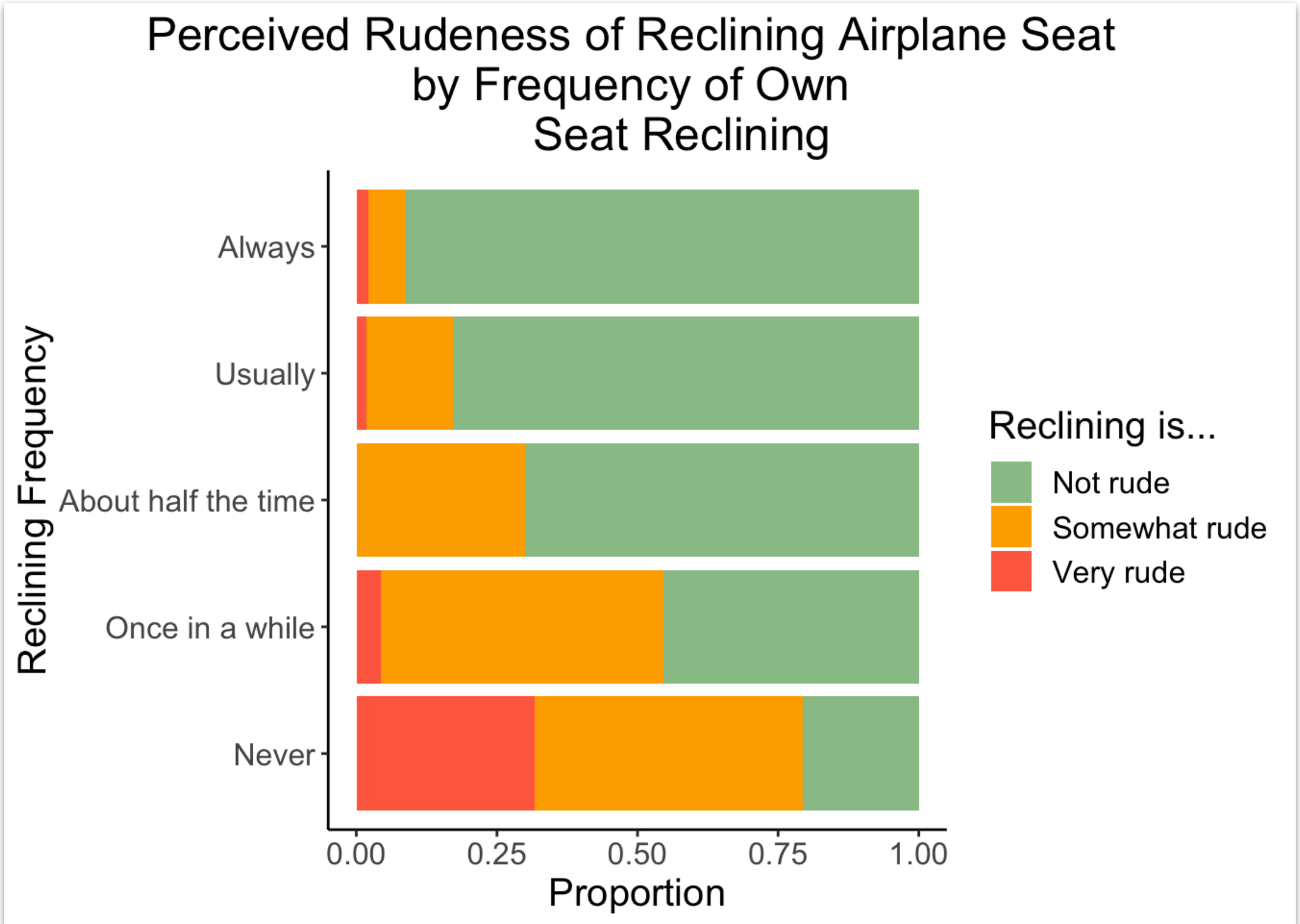

Alcohol Consumed per Person in 2010
100%
75%
50%
25%
0%
% of Total Servings
Africa
Asia
Europe
North America
Oceania
South America
Continent
Alcohol Type
beer
wine
spirit


Total Liters of Alcohol Consumed per Person in 2010
0
14.4

Homework


Children of parents with higher education achieve better test scores
66.4
73.1
0.025
0.020
0.015
0.010
0.005
0.000
Density
Parental Level of Education
some high school
some higher ed
0
25
50
75
100
Test Score (Reading + Writing + Math)

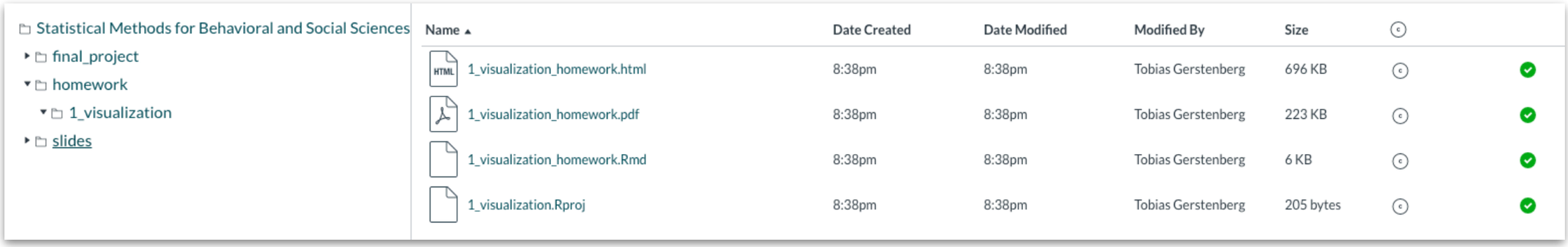
Homework


Perceived Rudeness of Reclining Airplane Seat
by Frequency of Own
Seat Reclining
Reclining Frequency
Always
Usually
About half the time
Once in a while
Never
0.00
0.25
0.50
0.75
1.00
Proportion
Reclining is...
Not rude
Somewhat rude
Very rude

Homework

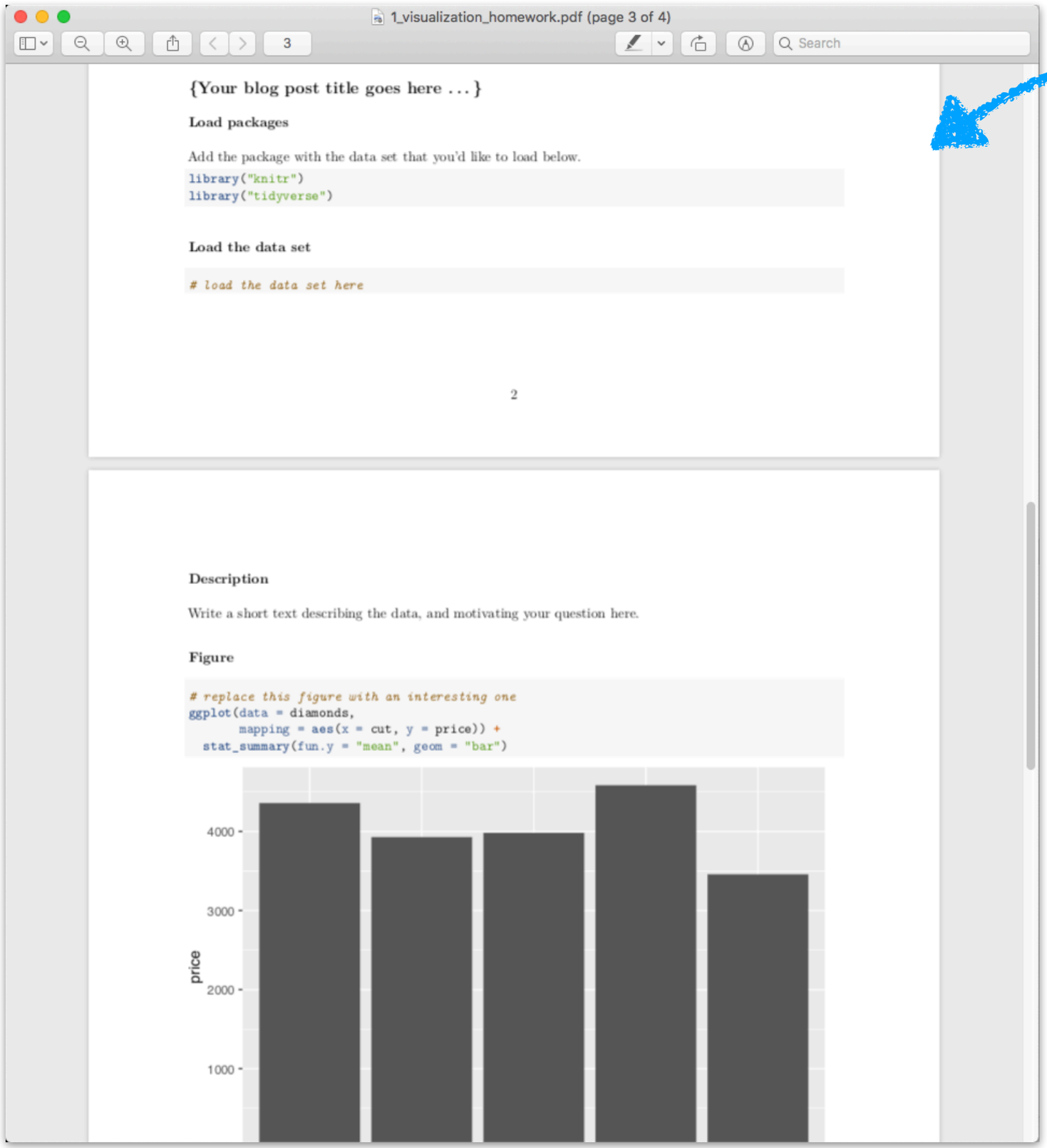
Homework is due by **Thursday 16th, 8pm**

Remember the 0 points for late submissions ...



Submit **one pdf file** (knitted with RMarkdown) that contains the code as well as the figure.

Homework



should look sort of like this ...

Homework

- install `tinytex` (https://yihui.name/tinytex/r/)
  - open `1-visualization.Rproj`
  - open `1-visualization_homework.Rmd` within RStudio

```
### Install tinytex

In order to knit an RMarkdown document to a pdf file, you have to install LaTeX on your computer. The
easiest way of doing so is via the `tinytex` package. Run the code in the following code chunk to do so:

```{r, eval=F}
install.packages("tinytex")
tinytex::install_tinytex()

# If you experience an error like the following when trying to knit to pdf:
# !LaTeX Error: File `xcolor.sty' not found.
# then run the following command: tinytex::tlmgr_install("xcolor")
# and try to knit again.
```

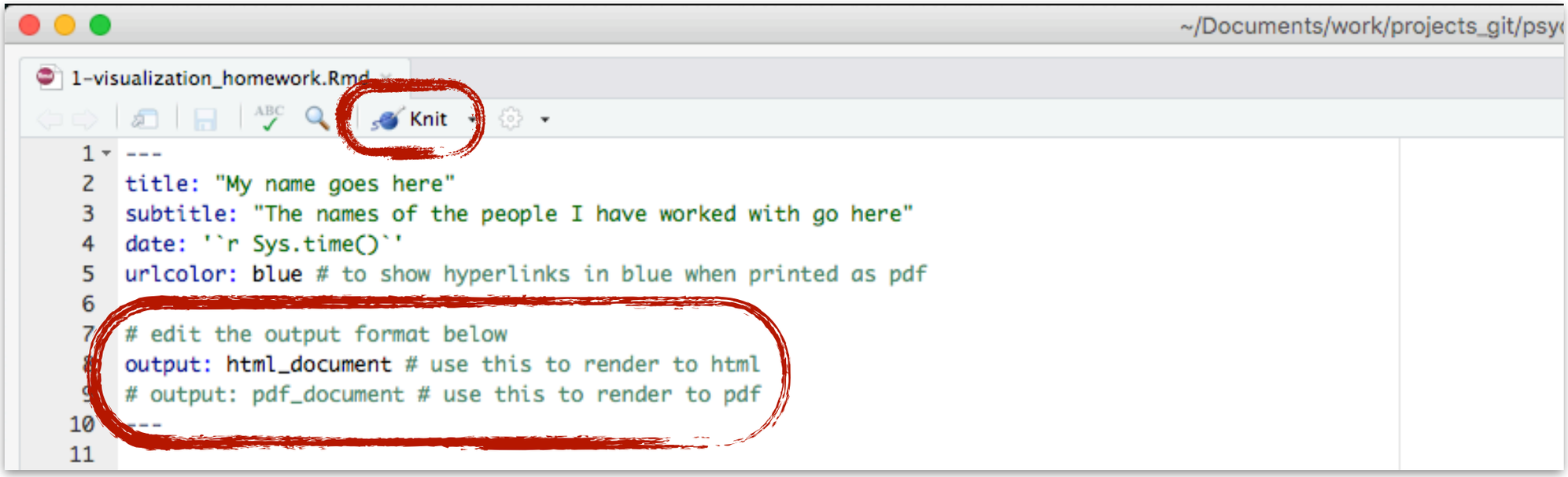
You can find out more about the `tinytex` package [here](https://yihui.org/tinytex/).
```

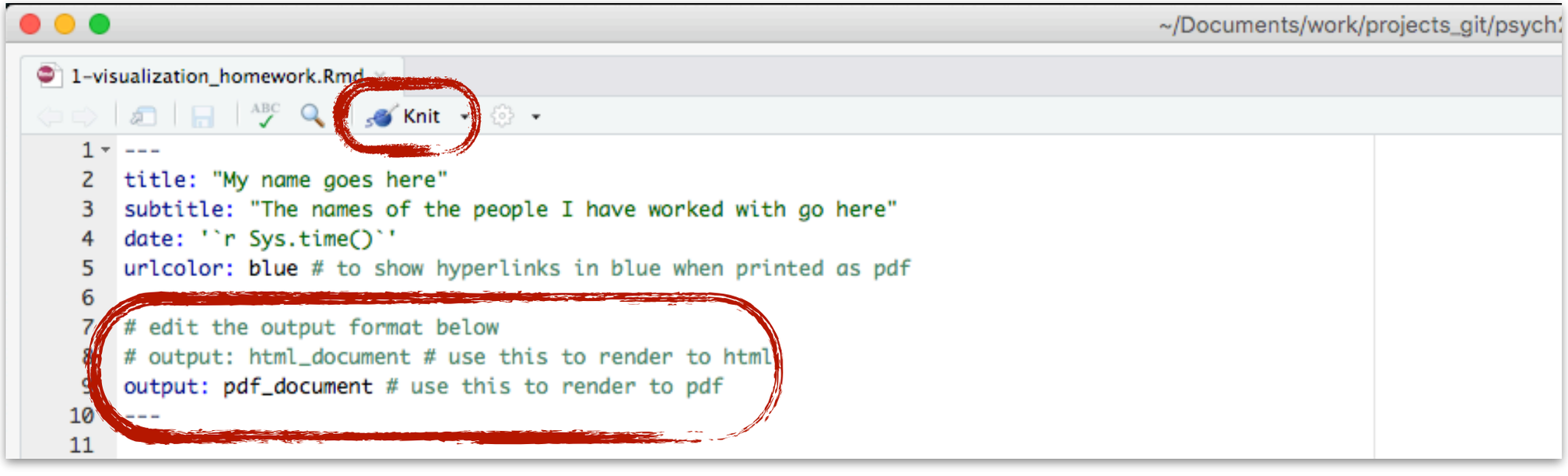
run this code

post on Piazza if you have any trouble getting this to work

Homework

- you can change the output format from html to pdf like so ...



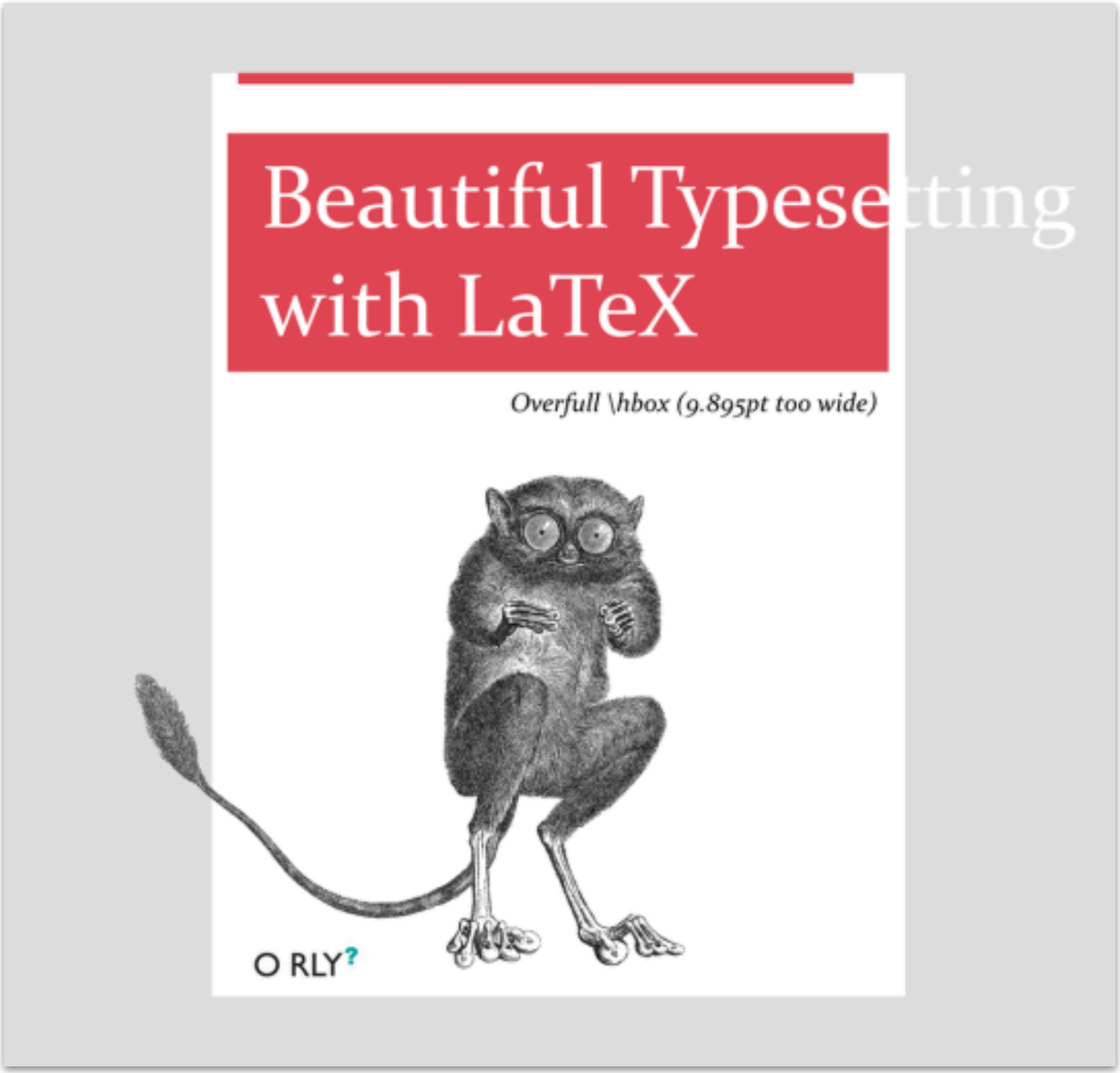


Homework

very long code without line break

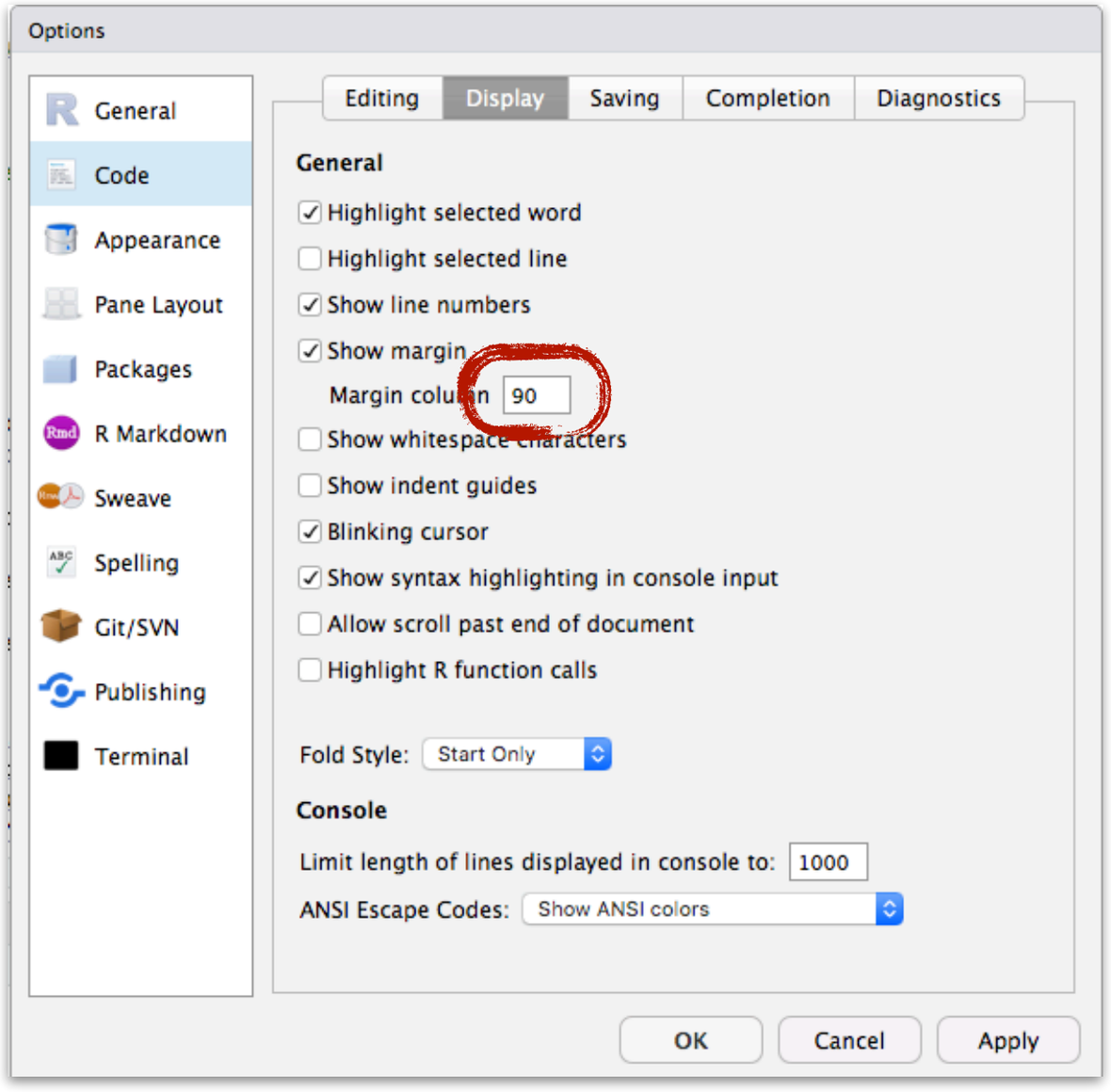
```
1 ggplot(data = df.diamonds, mapping = aes(y = price, x = color, fill = color, group = cut, shape = cut, ...)) +
2   stat_summary(fun.y = "mean", geom = "bar", color = "black") +
3   stat_summary(fun.data = "mean_cl_boot", geom = "linerange") +
4   facet_grid(rows = vars(cut), cols = vars(clarity))
```

Homework


Beautiful Typesetting with LaTeX
Overfull \hbox (9.895pt too wide)
O RLY?

Homework

- set the margin to 90 (and make sure not to go over that margin in code blocks)
- Preferences... > Code > Display



Homework

visualization2.Rmd

Knit | Insert | Run

```
---
title: "Class 3"
author: "Tobias Gerstenberg"
date: "January 11th, 2019"
output:
  bookdown::html_document2:
    toc: true
    toc_depth: 4
    theme: cosmo
    highlight: tango
---

```{r setup, include=FALSE}
# these options here change the formatting of how comments are rendered
knitr::opts_chunk$set(
  collapse = TRUE,
  comment = "#>")
```

Visualization 2

In this lecture, we will lift our `ggplot2` skills to the next level!

Learning objectives

- Deciding what plot is appropriate for what kind of data.
- Customizing plots: Take a sad plot and make it better.
- Saving plots.
- Making figure panels.
- Debugging.
- Making animations.
- Defining snippets.
```

margin column

12:1 (Top Level) R Markdown

Homework

- set the margin to 90 (and make sure not to go over that margin in code blocks)
- Preferences... > Code > Display

```
# take a look at the data sets that come with the package
data(package = "fivethirtyeight")

# take a look at the help file to get more information about the different data sets (not all packages
help("fivethirtyeight")

# the "fivethirtyeight" provides a detailed overview over the different data sets with this command
vignette("fivethirtyeight", package = "fivethirtyeight")

# to load a particular data set (e.g. US_births_2000_2014, replace with the name of the data set you'd
df.data = US_births_2000_2014
```

not good

only important in code chunks!

```
# take a look at the data sets that come with the package
data(package = "fivethirtyeight")

# take a look at the help file to get more information about the different data sets (not
# all packages have help files)
help("fivethirtyeight")

# the "fivethirtyeight" provides a detailed overview over the different data sets with
# this command
vignette("fivethirtyeight", package = "fivethirtyeight")

# to load a particular data set (e.g. US_births_2000_2014, replace with the name of the
# data set you'd liked to load) into your environment, run the following
df.data = US_births_2000_2014
```

good!

Some tips and tricks

Piazza

Actions

Practice Plot 3

Hi everyone,

I am trying to recreate the plot as part of Practice Plot 3 in Visualization 1.

this is great!

I wrote this so far:

```
ggplot(df.diamonds,
       aes(x = color,
           y = price,
           group = clarity,
           color = clarity))+
  stat_summary(fun.y = "mean",
               geom = "line") +
  stat_summary(fun.data = "mean_cl_boot",
               geom = "linerange")
```

The thickness of my lines is not right and I played a lot with "size = [number]" and managed to do a lot of weird graphs but didn't manage to recreate the actual plot.

Can someone tell me where and how to tell R that I want thicker lines?

Thanks!

rstudio

Piazza

this is even better!

- best way to get help is by posting a **reprex**
- **reprex =** reproducible example

reprex

CRAN 0.2.1 | build passing | build passing | codecov 78% | lifecycle stable

Overview

Prepare reprexes for posting to GitHub issues, StackOverflow, or Slack snippets. What is a `reprex`? It's a **repr**oducible **ex**ample, as coined by Romain Francois.

Given R code on the clipboard, selected in RStudio, as an expression (quoted or not), or in a file ...

- run it via `rmarkdown::render()`,
- with deliberate choices re: arguments and setup chunk.

Get resulting runnable code + output as

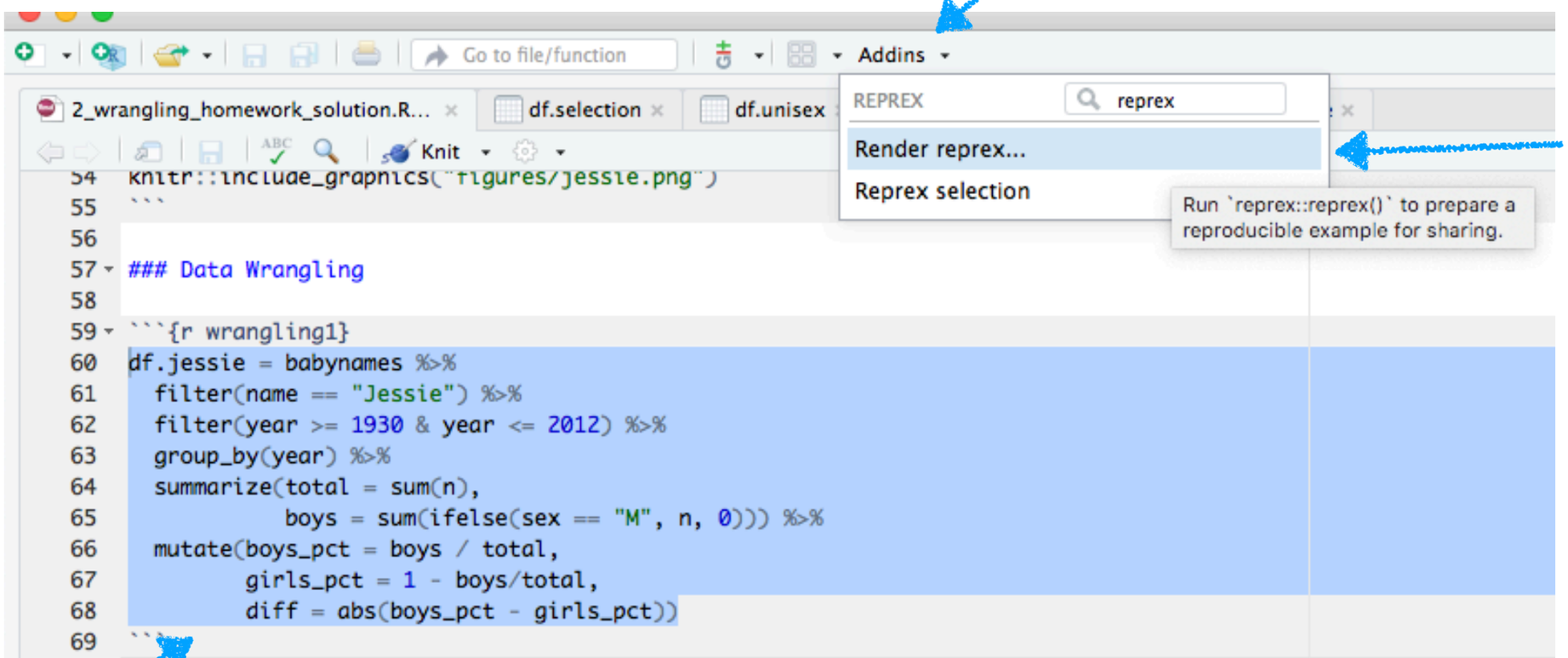
- Markdown, formatted for target venue, e.g. `gh` or `so`, or as
- R code, augmented with commented output.

Result is returned invisibly, placed on the clipboard, and written to a file. Preview an HTML version in RStudio viewer or default browser.

Piazza

```
install.package("reprex")
```

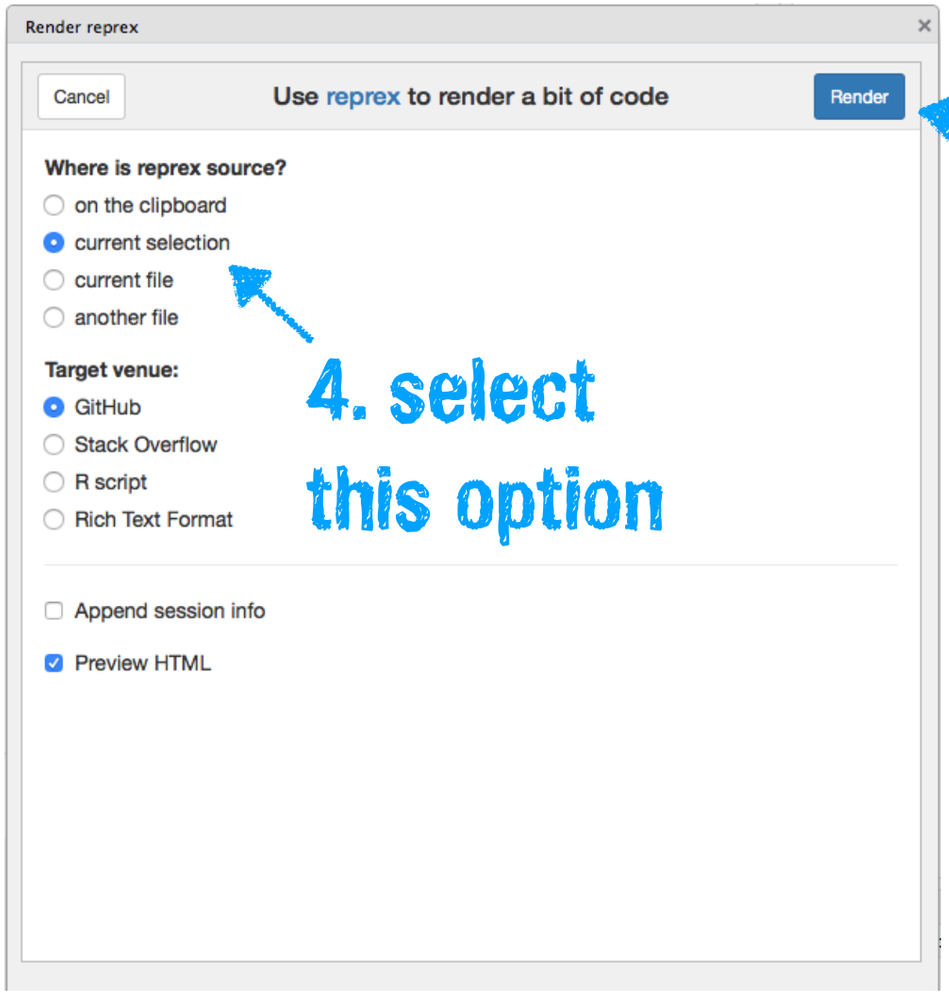
2. click on Addins



3. Render reprex

1. select the text

Piazza



6. copy and paste from the viewer

Piazza

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish

```
library("babynames")
library("tidyverse")
df.jessie = babynames %>%
  filter(name == "Jessie") %>%
  filter(year >= 1930 & year <= 2012) %>%
  group_by(year) %>%
  summarize(total = sum(n),
            boys = sum(ifelse(sex == "M", n, 0))) %>%
  mutate(boys_pct = boys / total,
         girls_pct = 1 - boys/total,
         diff = abs(boys_pct - girls_pct)) %>%
  print()
#> # A tibble: 83 x 6
#>     year total  boys boys_pct girls_pct  diff
#>    <dbl> <int> <dbl>    <dbl>     <dbl> <dbl>
#>  1  1930  3525  1329    0.377     0.623 0.246
#>  2  1931  3196  1267    0.396     0.604 0.207
#>  3  1932  3178  1282    0.403     0.597 0.193
#>  4  1933  2886  1079    0.374     0.626 0.252
#>  5  1934  2883  1090    0.378     0.622 0.244
#>  6  1935  2721  1103    0.405     0.595 0.189
#>  7  1936  2599  1012    0.389     0.611 0.221
#>  8  1937  2589  1041    0.402     0.598 0.196
#>  9  1938  2446   970    0.397     0.603 0.207
#> 10  1939  2454  1058    0.431     0.569 0.138
#> # … with 73 more rows
```

7. make sure to load necessary packages in your reprex

Created on 2019-01-24 by the reprex package (v0.2.1)

Logistics

Coding

blue

pink

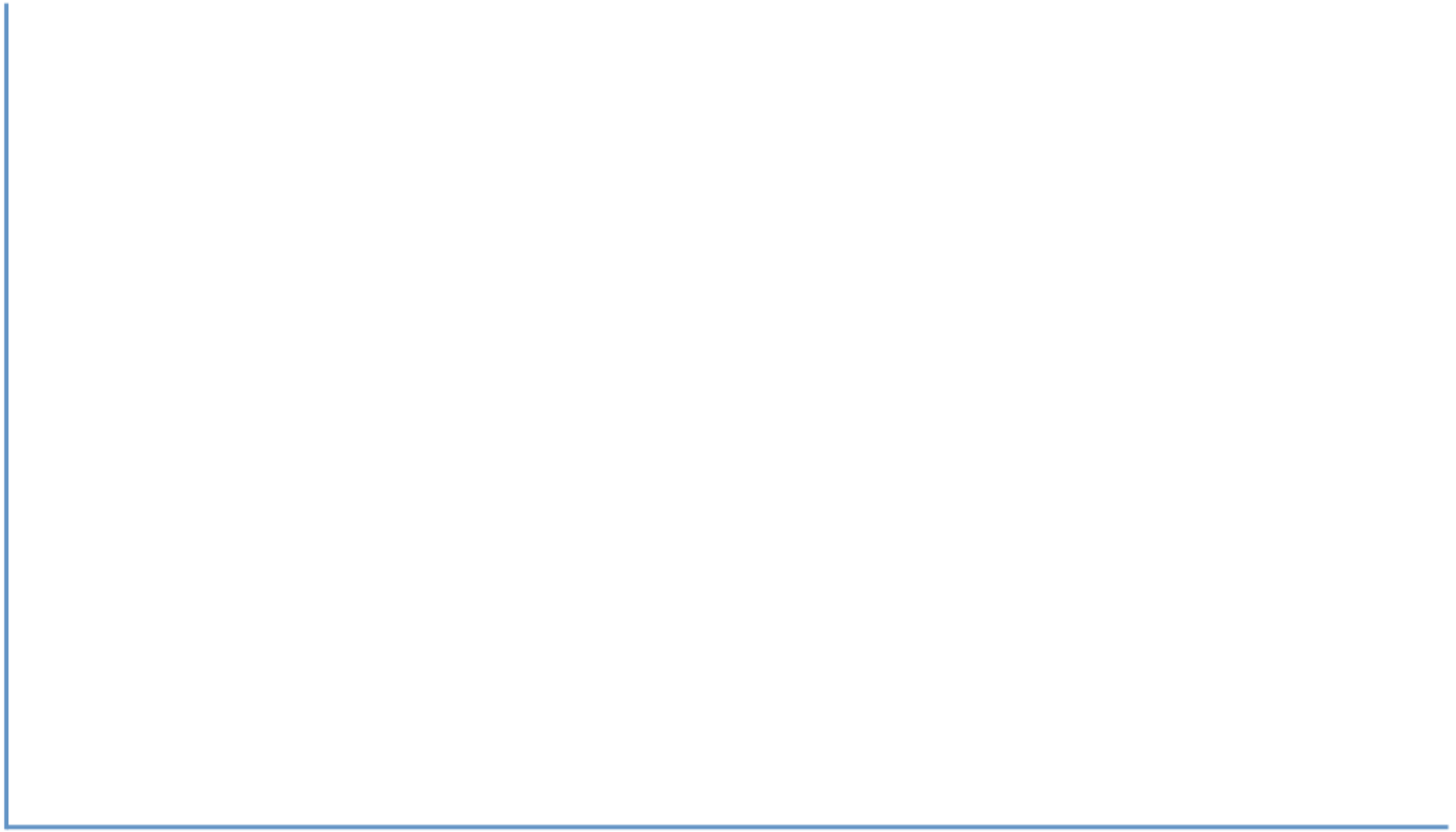
RStudio & visualization time!

blue

pink

Feedback

How was the pace of today's class?



How happy were you with today's class overall?

What did you like about today's class? What could be improved next time?

Thank you!