

Visualization 2



Things that came up ...

10 guidelines for effective data visualization (Keheller & Wagner, 2011)

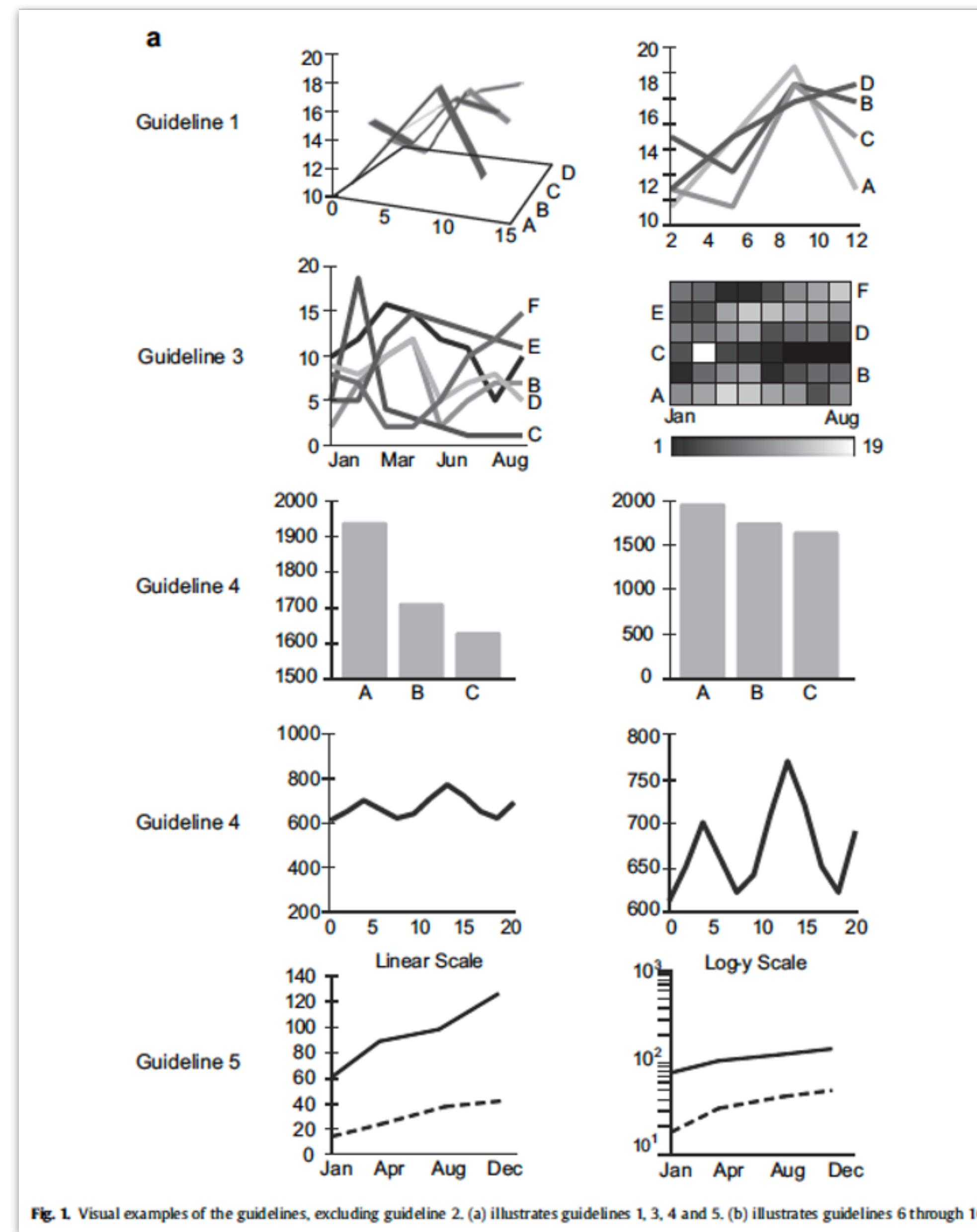


Fig. 1. Visual examples of the guidelines, excluding guideline 2. (a) illustrates guidelines 1, 3, 4 and 5.

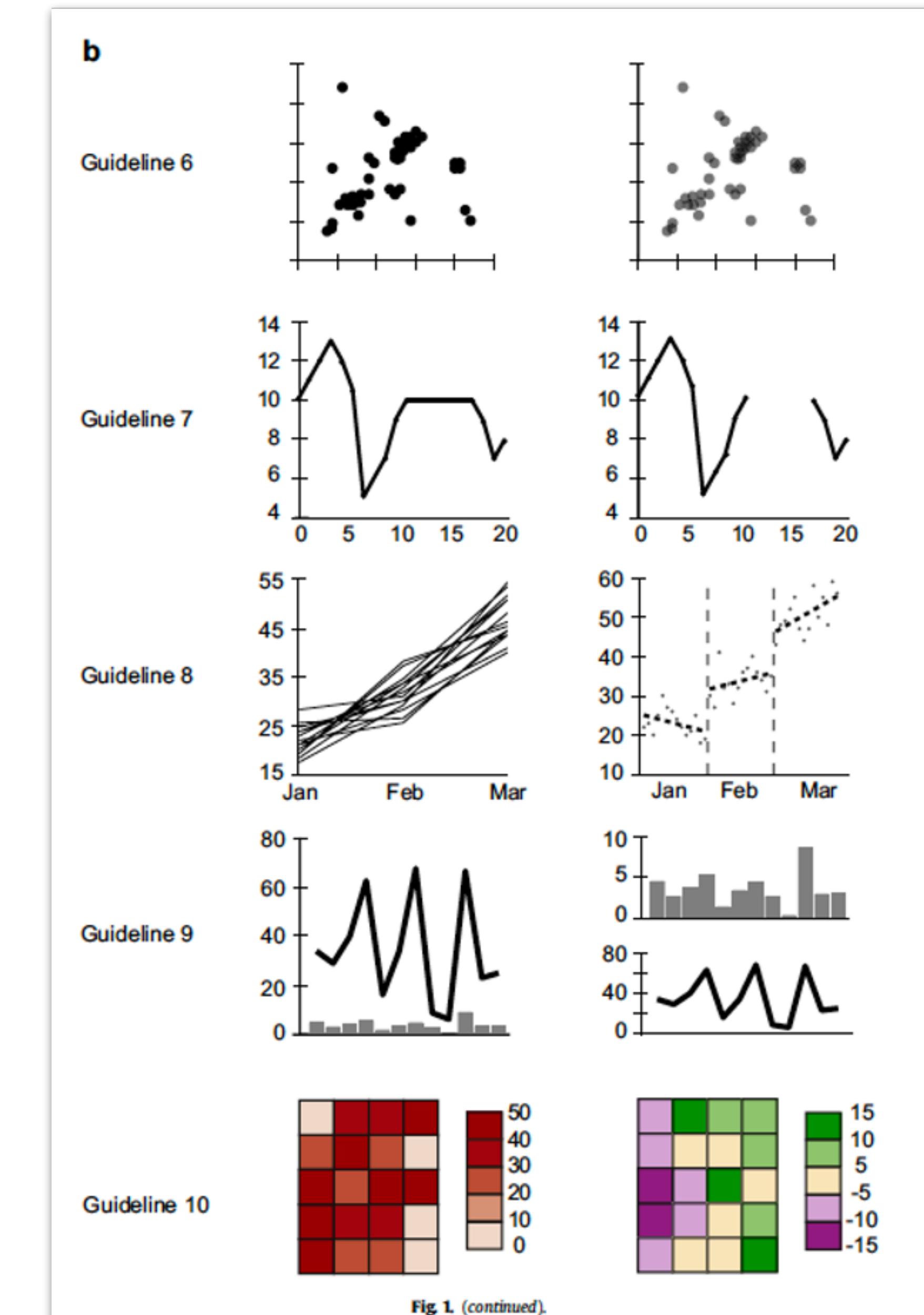


Fig. 1. (continued).

Nice interactive visualization

The **React** Graph Gallery

LEARN ALL CHART TYPES ABOUT SUBSCRIBE

Why R^2 Alone Fails



R^2 and **correlation** are often seen as definitive measures to validate the relationship between two variables.

This post features an interactive sandbox that explores several edge cases, demonstrating how relying on these summary statistics without visualizing the data can be **dangerously misleading**.

Useful links

[inspiration](#) [d3 gallery](#) [About this chart](#)

<https://www.react-graph-gallery.com/example/scatterplot-r2-playground>

Homework

Homework

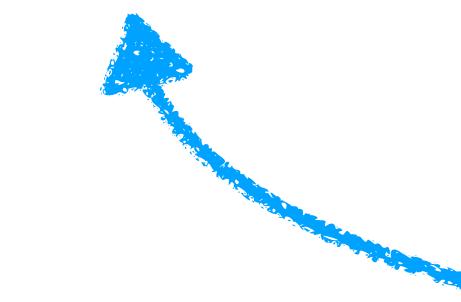
In this homework, **you'll write a short blog post** about a data set. Your goal is to tell us something interesting using a well-crafted, thoughtfully-prepared data graphic.

Grading Rubric

There are 15 possible points for this homework.

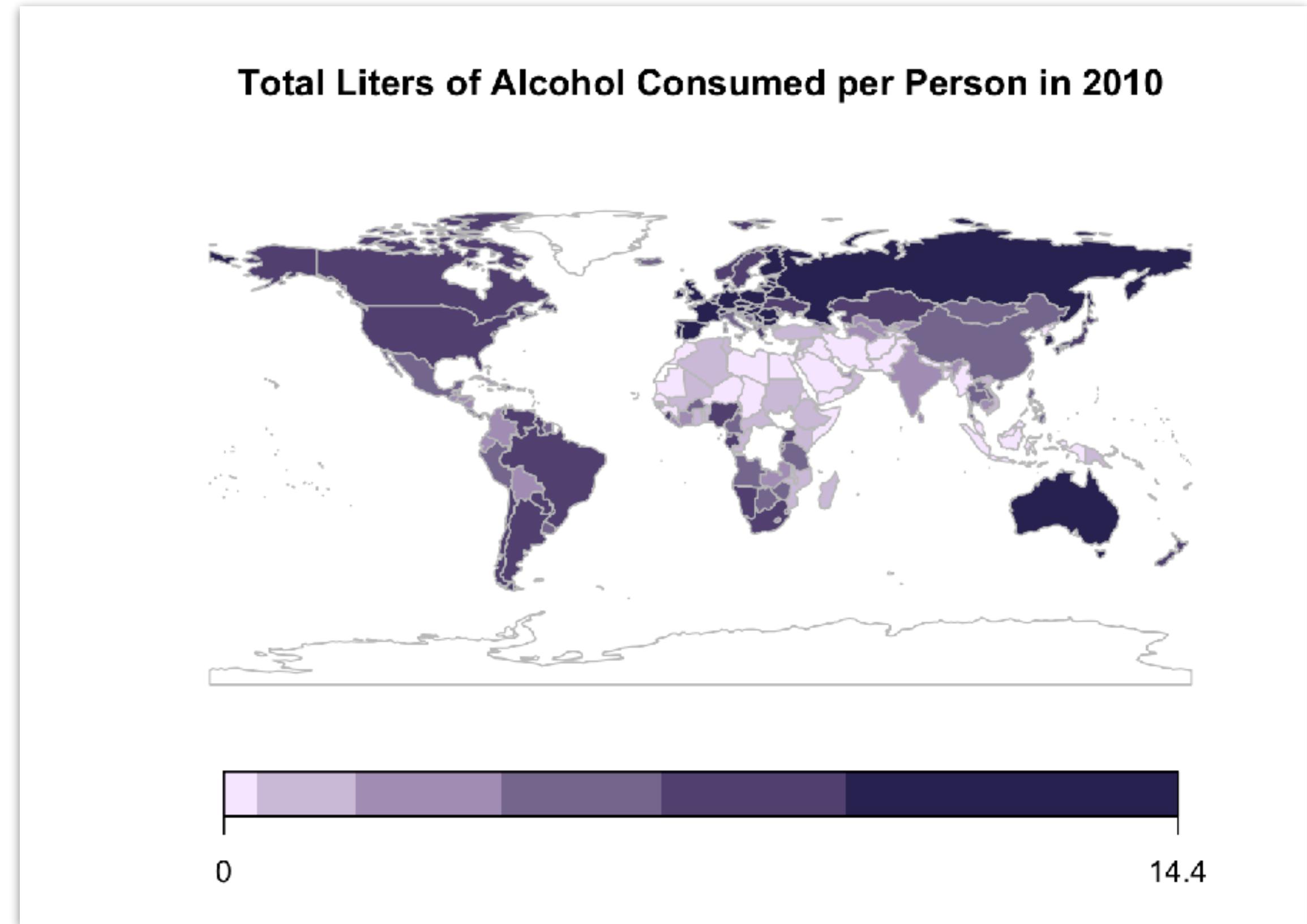
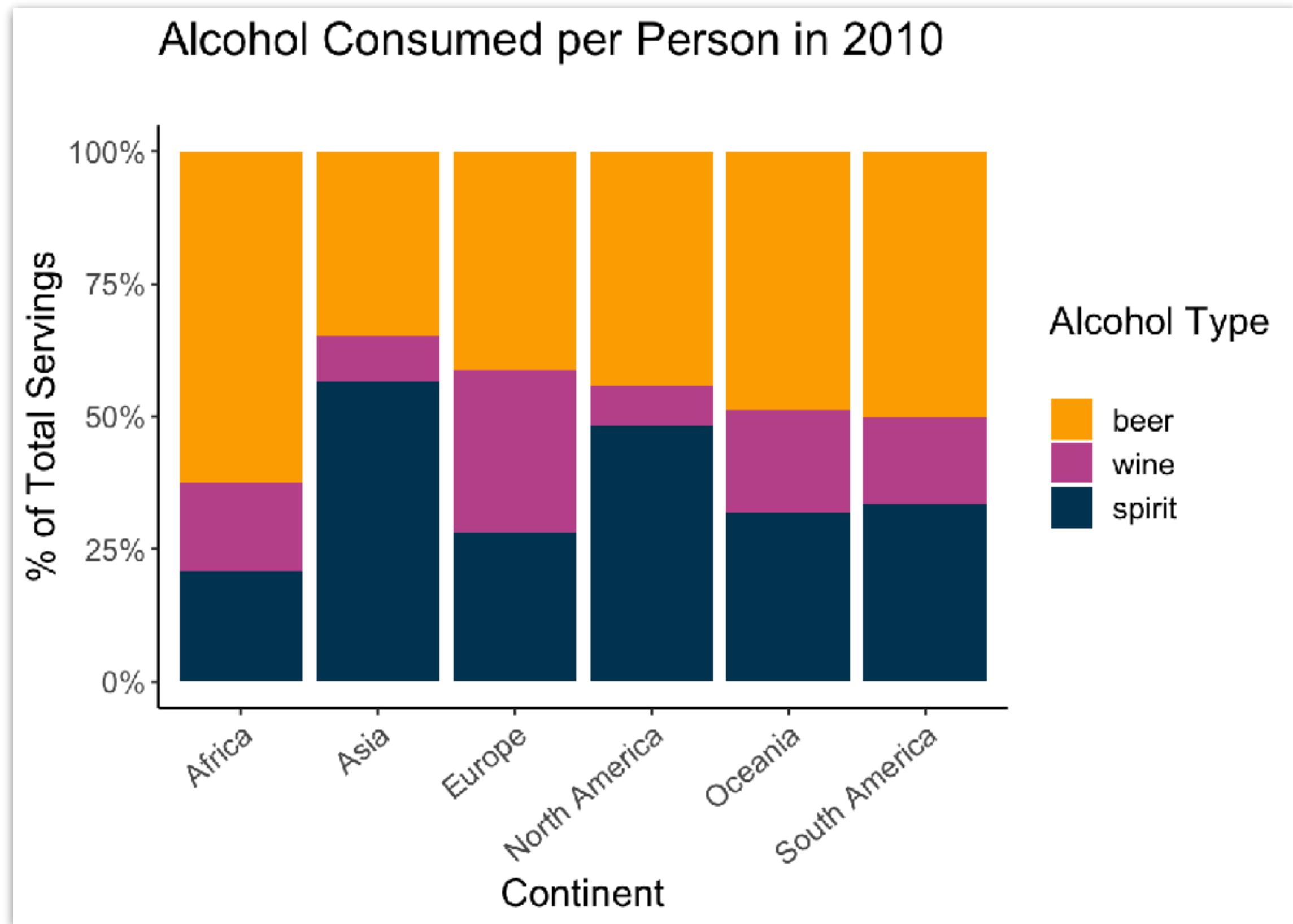
Here are some of the things we care about:

- include all the code that you used to generate the plot (3 points)
- consistent coding style (2 points)
- all the code can be seen in the knitted pdf document (1 point)
- an interesting plot that demonstrates what you've learned in class (4 points)
- a figure caption that is sufficient to understand the plot (2 points)
- a succinct blog post to go with the plot (3 points)

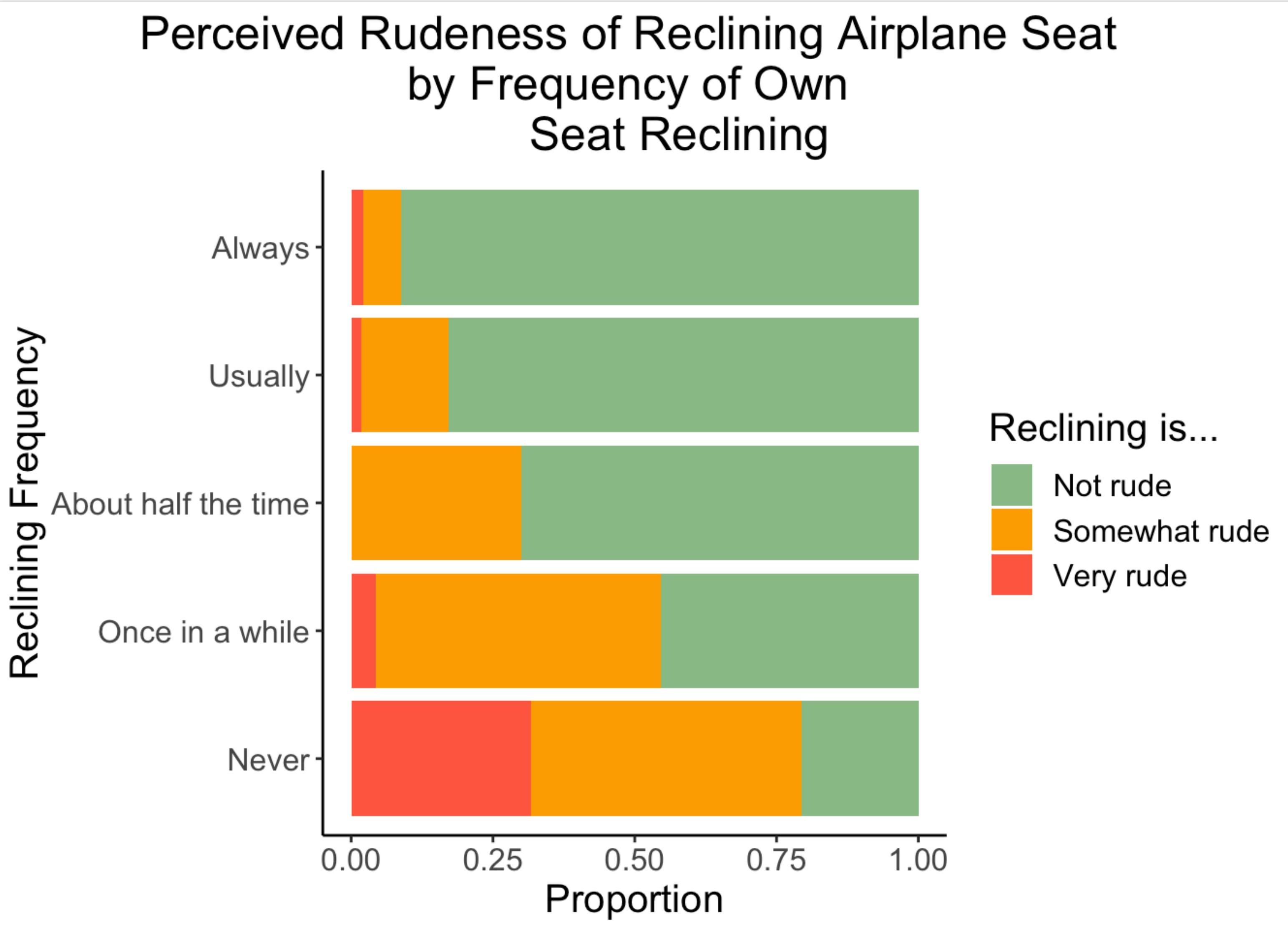


is at the bottom of the RMarkdown file

Homework



Homework

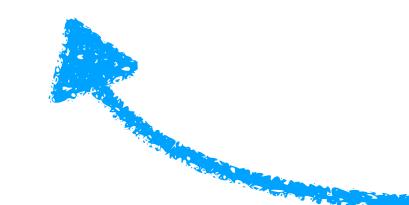
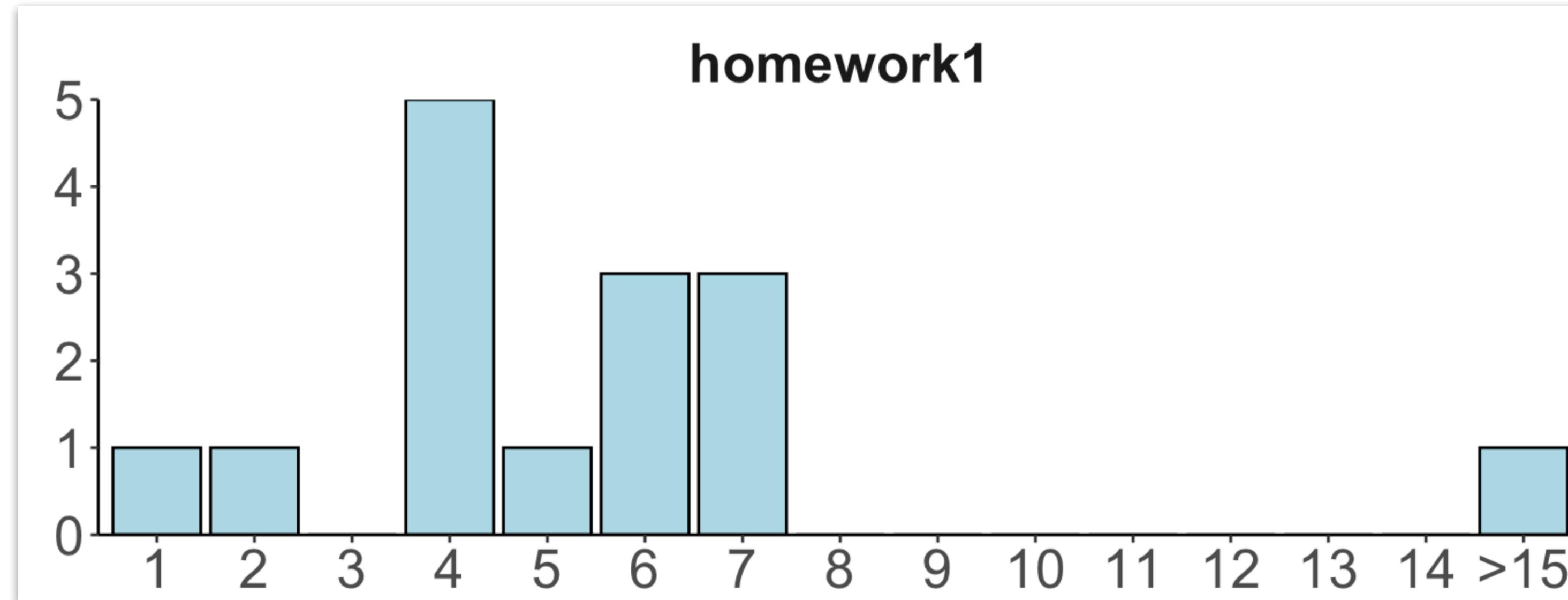


Homework

Homework is due by **Thursday 15th, 8pm**

Submit **two files**:

- the raw **.Rmd** file
- the rendered **.pdf** file that contains the code as well as the figure.



how long it took people in the past

Homework

1_visualization_homework.pdf (page 3 of 4)

{Your blog post title goes here ...}

Load packages

Add the package with the data set that you'd like to load below.

```
library("knitr")
library("tidyverse")
```

Load the data set

```
# load the data set here
```

2

Description

Write a short text describing the data, and motivating your question here.

Figure

```
# replace this figure with an interesting one
ggplot(data = diamonds,
       mapping = aes(x = cut, y = price)) +
  stat_summary(fun.y = "mean", geom = "bar")
```

Cut	Mean Price (approx.)
Good	4200
Poor	3900
Very Good	4000
Very Good	4400
Fair	3500

should look sort of
like this ...

Homework

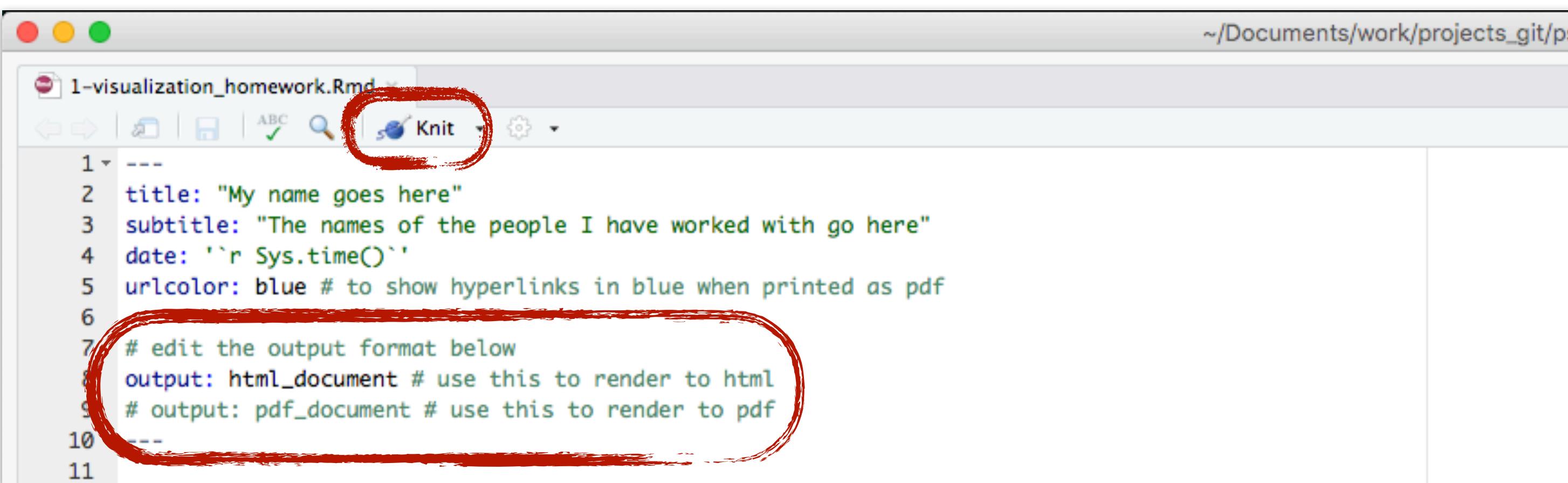
- install `tinytex` (<https://yihui.name/tinytex/r/>)
 - open `1-visualization.Rproj`
 - open `1-visualization_homework.Rmd` within RStudio

```
30 ## Install tinytex
31
32 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:
33
34 ```{r eval=FALSE}
35 install.packages("tinytex")
36 tinytex::install_tinytex()
37
38 # If you experience an error like the following when trying to knit to pdf:
39 # !LaTeX Error: File 'xcolor.sty' not found.
40 # then run the following command: tinytex::tlmgr_install("xcolor")
41 # and try to knit again.
42 ```
43
44 You can find out more about the `tinytex` package \[here\](https://yihui.org/tinytex/).
```

run this code

Homework

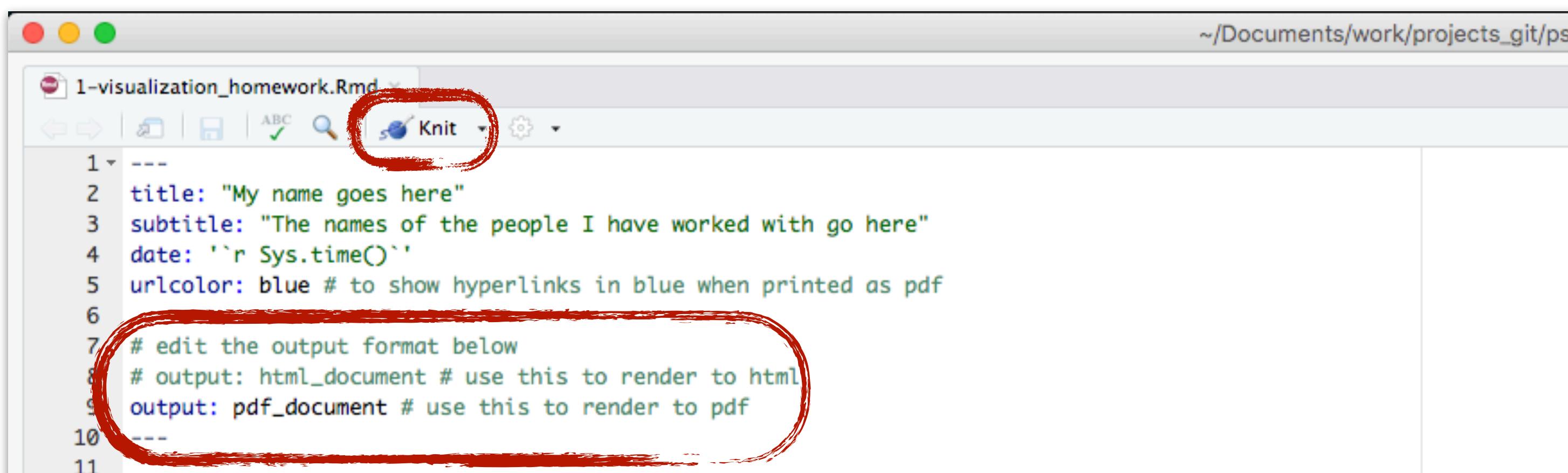
- you can change the output format from html to pdf like so ...



A screenshot of the RStudio interface showing an R Markdown file named "1-visualization_homework.Rmd". The code in the editor is:

```
1 ---  
2 title: "My name goes here"  
3 subtitle: "The names of the people I have worked with go here"  
4 date: ``r Sys.time()``  
5 urlcolor: blue # to show hyperlinks in blue when printed as pdf  
6  
7 # edit the output format below  
8 output: html_document # use this to render to html  
9 # output: pdf_document # use this to render to pdf  
10 ---  
11
```

The line "# edit the output format below" and the subsequent lines "output: html_document" and "output: pdf_document" are circled in red.



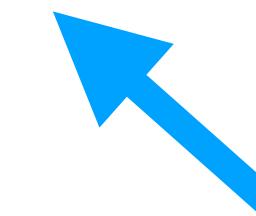
A screenshot of the RStudio interface showing the same R Markdown file "1-visualization_homework.Rmd". The code in the editor is identical to the previous screenshot:

```
1 ---  
2 title: "My name goes here"  
3 subtitle: "The names of the people I have worked with go here"  
4 date: ``r Sys.time()``  
5 urlcolor: blue # to show hyperlinks in blue when printed as pdf  
6  
7 # edit the output format below  
8 # output: html_document # use this to render to html  
9 output: pdf_document # use this to render to pdf  
10 ---  
11
```

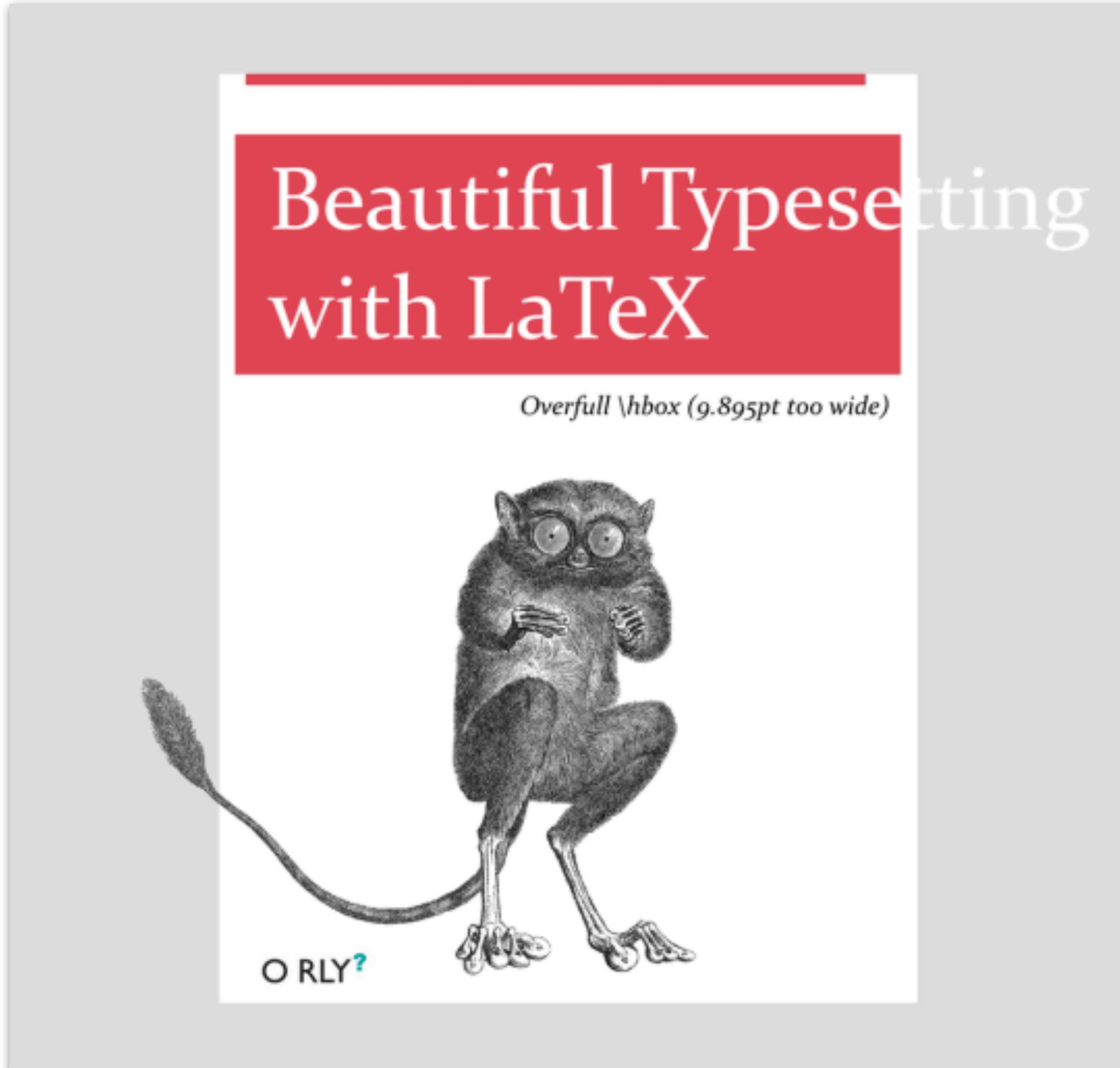
The line "# edit the output format below" and the subsequent lines "output: html_document" and "output: pdf_document" are circled in red.

Homework

```
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))
```

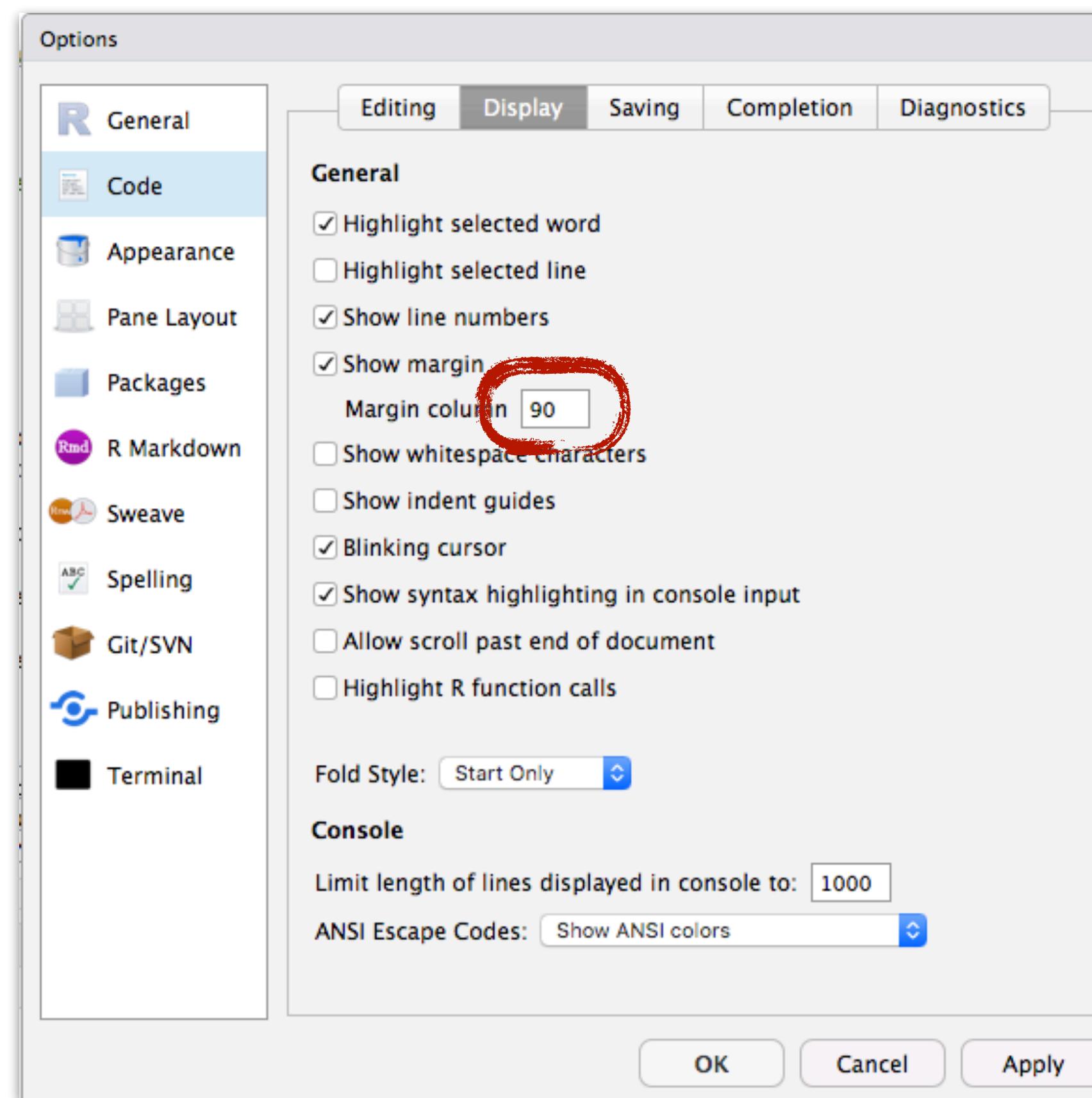


very long code without line break

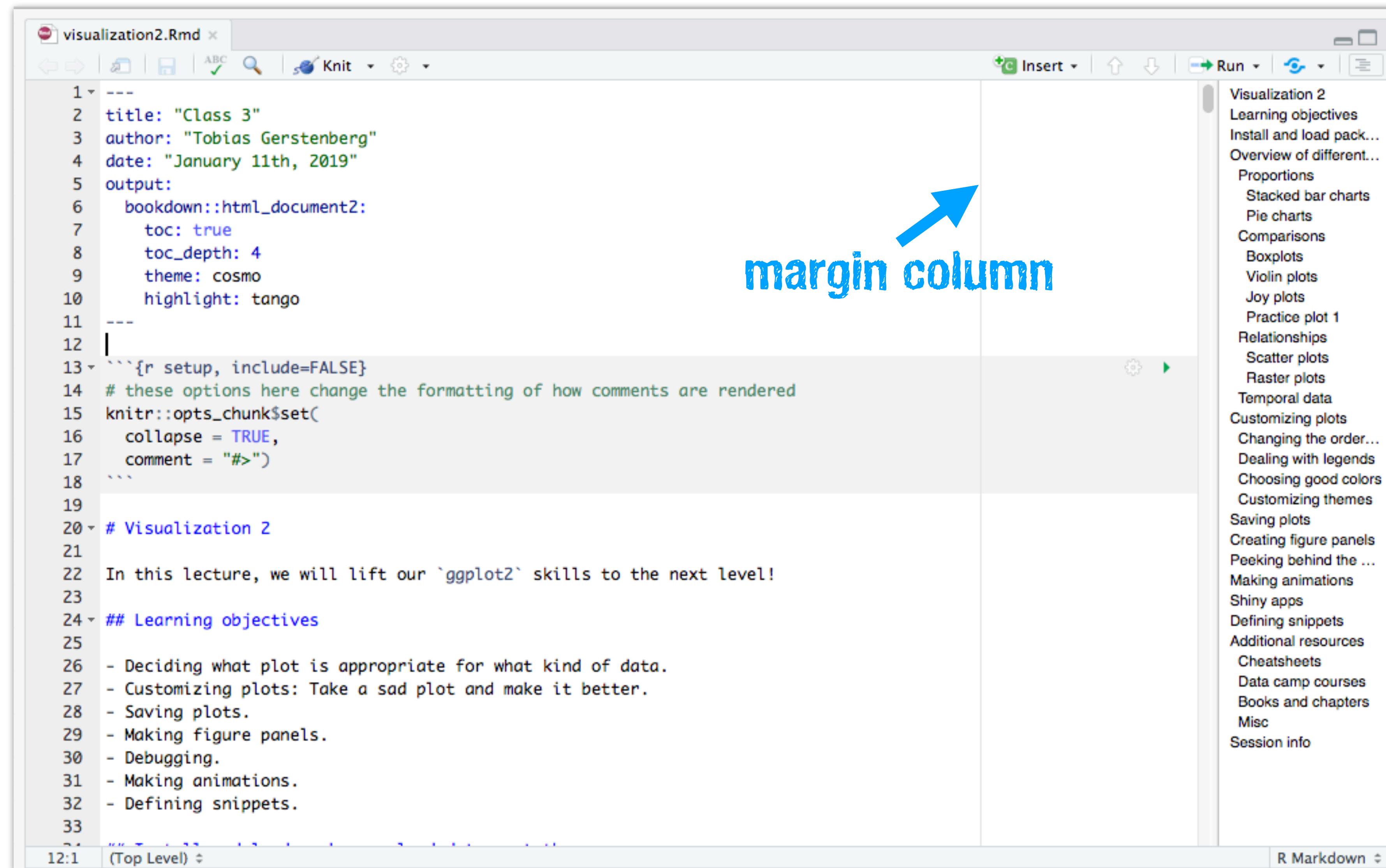


Homework

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



Homework



margin column

```
1 ---  
2 title: "Class 3"  
3 author: "Tobias Gerstenberg"  
4 date: "January 11th, 2019"  
5 output:  
6   bookdown::html_document2:  
7     toc: true  
8     toc_depth: 4  
9     theme: cosmo  
10    highlight: tango  
11 ---  
12 |  
13 ```{r setup, include=FALSE}  
14 # these options here change the formatting of how comments are rendered  
15 knitr::opts_chunk$set(  
16   collapse = TRUE,  
17   comment = "#>")  
18 ````  
19  
20 # Visualization 2  
21  
22 In this lecture, we will lift our `ggplot2` skills to the next level!  
23  
24 ## Learning objectives  
25  
26 - Deciding what plot is appropriate for what kind of data.  
27 - Customizing plots: Take a sad plot and make it better.  
28 - Saving plots.  
29 - Making figure panels.  
30 - Debugging.  
31 - Making animations.  
32 - Defining snippets.  
33  
34
```

12.1 (Top Level) R Markdown

Visualization 2
Learning objectives
Install and load pack...
Overview of different...
Proportions
Stacked bar charts
Pie charts
Comparisons
Boxplots
Violin plots
Joy plots
Practice plot 1
Relationships
Scatter plots
Raster plots
Temporal data
Customizing plots
Changing the order...
Dealing with legends
Choosing good colors
Customizing themes
Saving plots
Creating figure panels
Peeking behind the ...
Making animations
Shiny apps
Defining snippets
Additional resources
Cheatsheets
Data camp courses
Books and chapters
Misc
Session info

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

```
# 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
```

only important in
code chunks!

good!

Homework

The screenshot shows the Ed Discussion interface for the course psych252. The left sidebar lists categories: COURSES (psych252), CATEGORIES (General, R/RStudio, Lectures, Homework, HW1, HW2, HW3, HW4, HW5, HW6, HW7, Midterm, Final Project, Social, Random). The main area displays a "No threads" message with a "Be the first to create a thread!" button. A search bar and a "Filter" dropdown are present. On the right, a "New Post" dialog is open, showing tabs for Question, Post (selected), and Announcement. The "Post" tab has fields for Title, Category (Homework selected), Subcategory (HW1 selected), and a rich text editor toolbar. Below the toolbar are checkboxes for Pinned (Keep at top of thread list), Private (Visible to you and staff only), Anonymous (Hide your name from students), Anonymous Comments (Allow anonymous comments), and Megathread (Resolvable comments). A "Post" button is at the bottom right.

post on Ed Discussion if you have any questions
about the homework

RStudio & visualization time!



Please help.

blue

I'm done.

pink

Anatomy of a nice ggplot

```
1 # ggplot call with global aesthetics  
2 ggplot(data = data,  
3         mapping = aes(x = cause,  
4                             y = effect)) +  
5 # add geometric objects (geoms)  
6 geom_point() +  
7 stat_summary(fun.y = "mean", geom = "point") +  
8 ... +  
9 # add text objects  
10 geom_text() +  
11 annotate() +  
12 # adjust axes and coordinates  
13 scale_x_continuous() +  
14 scale_y_continuous() +  
15 coord_cartesian() +  
16 # define plot title, and axis titles  
17 labs(title = "Title",  
18       x = "Cause",  
19       y = "Effect") +  
20 # change global aspects of the plot  
21 theme(text = element_text(size = 20),  
22        plot.margin = margin(t = 1, b = 1, l = 0.5, r = 0.5, unit = "cm")) +  
23 # save the plot  
24 ggsave(filename = "super_nice_plot.pdf",  
25         width = 8,  
26         height = 6)
```

The diagram illustrates the flow of a ggplot command with handwritten annotations:

- A blue arrow points from the word "what?" to the first line of code, "# ggplot call with global aesthetics".
- A blue arrow points from the word "how?" to the second line of code, "ggplot(data = data,".
- A blue arrow points from the words "add some text?" to the ninth line of code, "geom_text() +".
- A blue arrow points from the words "'local' adjustments" to the twelfth line of code, "scale_x_continuous() +".
- A blue arrow points from the words "'global' adjustments" to the twenty-first line of code, "theme(text = element_text(size = 20),".
- A blue arrow points from the words "save the beauty!" to the twenty-fourth line of code, "ggsave(filename = "super_nice_plot.pdf",".

Feedback

How was the pace of today's class?

much
too
slow

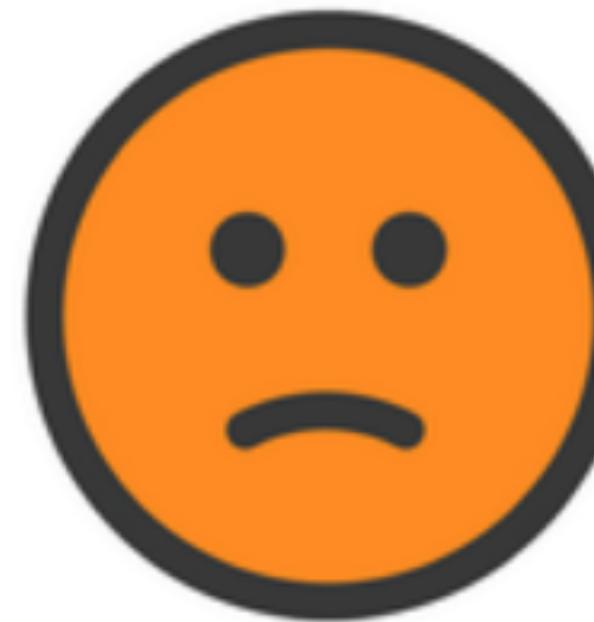
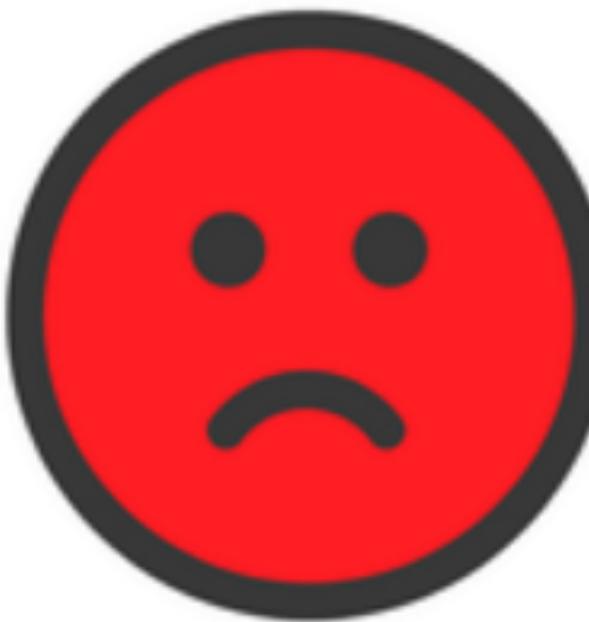
a little
too
slow

just
right

a little
too
fast

much
too
fast

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!