

SOC 4650/5650: Exercise 02b - Plotting and Reproducible Examples

Christopher Prener, Ph.D.

January 29nd, 2018

Directions

Use the same notebook from the prior exercise and continue adding to it!

Installing Dependencies in RStudio

1. Install reprex:

```
install.packages("reprex")
```

2. Install stlData if you have not done so already:

```
install.packages("devtools")  
devtools::install_github("chris-prener/stlData")
```

Loading Packages and Assigning Data to Objects

3. Open the Lecture Prep 02 notebook and add a fourth second-level heading that looks like this:

```
## Load Data
```

4. Add some introductory text on the next line:

```
The following code loads the data package and assigns  
our data to a data frame in our global environment:
```

5. Add a new code chunk named data-load that contains the following code and execute it:

```
```{r data-load}  
library(stlData) # data source
leadData <- stlLead
```
```

Viewing Data Objects

6. Create a spreadsheet-like view of your data by entering the following function into the *console*:

```
View(leadData)
```

Creating Your First Plot

7. In the Lecture Prep 02 notebook, add a fifth second-level heading followed by a third-level heading:

```
## Part 1
### Question 1
```

8. Add some introductory text on the next line that includes back-ticks around the word `pctElevated`:

```
The first question in this assignment asks for a
histogram of the 'pctElevated' variable.
```

9. Add a new code chunk named `data-load` that contains the following code and execute it:

```
```{r lead-plot-01}
ggplot() +
 geom_histogram(leadData, mapping = aes(pctElevated))
```
```

10. Add a interpretation to your narrative text below the plot:

```
We can see that the bulk of the census tracts have
elevated blood lead level rates of between three and
ten percent, but there are some tracts where nearly
a quarter of children have elevated rates.
```

Saving Plots

11. In the Lecture Prep 02 notebook, add a another third-level heading along with the following narrative:

```
### Question 2
Next, we want to save our plot using the 'ggsave()'
function:
```

12. Add a new subdirectory to your project named results.
13. Add a new code chunk named data-load that contains the following code and execute it:

```
```{r save-plot-01}
ggsave("results/leadHistogram.png")
```
```

Knitting, Committing, and Pushing

14. Save your document and be sure that the .nb.html file is updated.
15. **Knit** your document to produce the GitHub output.
16. Open GitHub Desktop, **commit** the change, and **push** it to GitHub.com
17. Open the LP-02 subdirectory on GitHub.com and check out how each of the different files appears on the website.

Reproducible Examples

18. In RStudio, go to File > New File > R Script
19. Copy the following syntax into the script from your notebook:

```
library(ggplot2) # creating plots
library(stlData) # data source

leadData <- stlLead

ggplot() +
  geom_histogram(leadData, mapping = aes(pctElevated))
```

20. Highlight this entire set of code, and copy it to your system's clipboard.
21. In the *console*, open the `reprex` package the execute the `reprex()` function:

```
library(reprex)
reprex()
```

22. Once the output appears in the viewer, open up Slack and create a direct message with both Brandon and Chris. In the message composition field, paste the reproducible example. Delete the `![](` before the link the the parenthesis afterward.
23. On the line with the initial set of backticks, delete the character `r`.
24. Above the initial set of backticks, compose a quick question inquiring about the warning message you received the `binwidth`.
25. Send the message!