Problem Set 1 - PPOL 6802

Alex Lundry

2025-01-23

Brief Explanation of R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. What's most helpful about it is the ease with which you can integrate formatted text, code and charts. For more details on using R Markdown see http://rmarkdown.rstudio.com. You can also find an extremely helpful "cheatsheet" here.

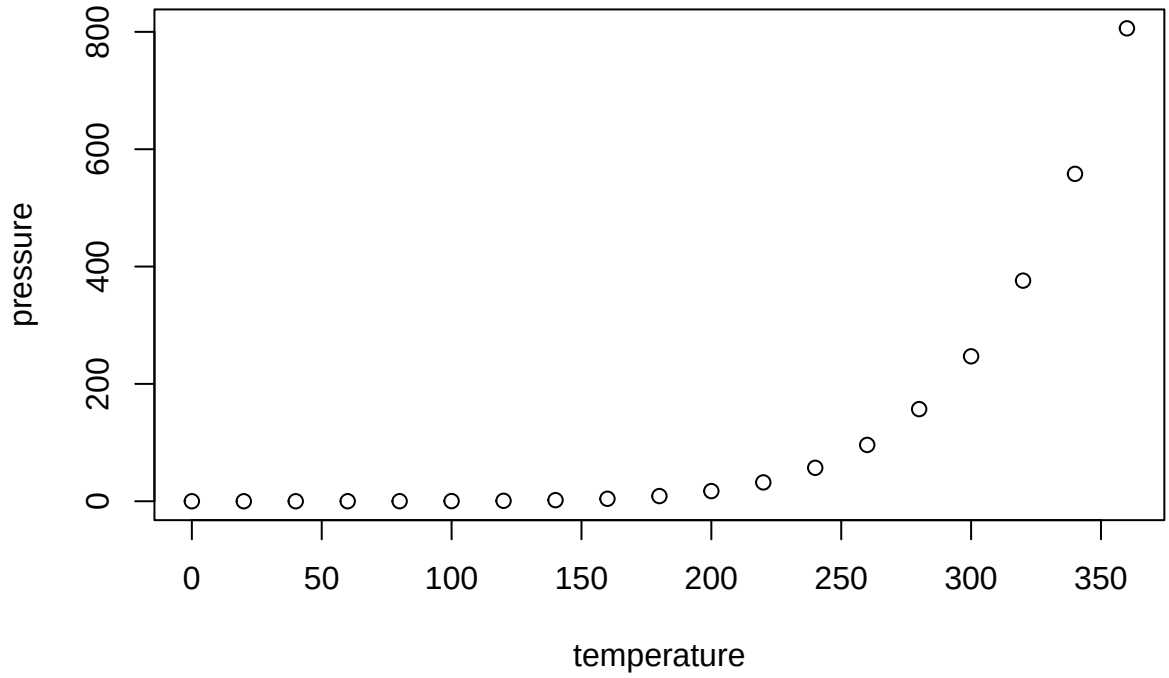
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot. For your problem sets you MUST show your code, so be sure that each code chunk for the problem set is set to `echo = TRUE`.

Problem Set #1

Here is where you will provide the answers to problem set 1. I will provide you with the basic "buckets" to put everything in to make the R Markdown creation process as straightforward as possible. But remember that if you run into any issues, ChatGPT can be extremely helpful!

Loading the Data

To complete this problem set, you will need to download the necessary data from Github. You can do that by running the code below:

```
library(tidyverse)

d1 <- read_csv("https://raw.githubusercontent.com/alexlundry/ppol6802_spring24/main/problem_sets/dataset

d2 <- read_csv("https://raw.githubusercontent.com/cphalpert/census-regions/master/us%20census%20bureau%2

d3 <- left_join(d1, d2, by = "State")
```

Question 1 - Tidyverse

You've been asked to identify the average 2020 Trump % county margin in the South within counties that are majority-minority and those that are not. Write Tidyverse syntax that does the following:

- Select only the following variables, making sure they are in this order: NAME, State, Region, Division, pct_white, and any variables with 2020 in the name.
- Filter so that we only have counties in the "South" region.
- Create a new binary variable that indicates whether a county is a majority-minority county; that is, the White % is less than 50%.
- Create a new variable that is the percent margin between the Republican and Democratic vote share.
- Group by your new majority-minority variable.
- Summarize the data to find the mean percent margin.

Below you will see the correct output. Show and run tidyverse code that will match this output.

```
## # A tibble: 2 x 2
##   maj_minority mean_pct_margin
##   <lgl>                  <dbl>
## 1 FALSE                  0.407
## 2 TRUE                  -0.208
```

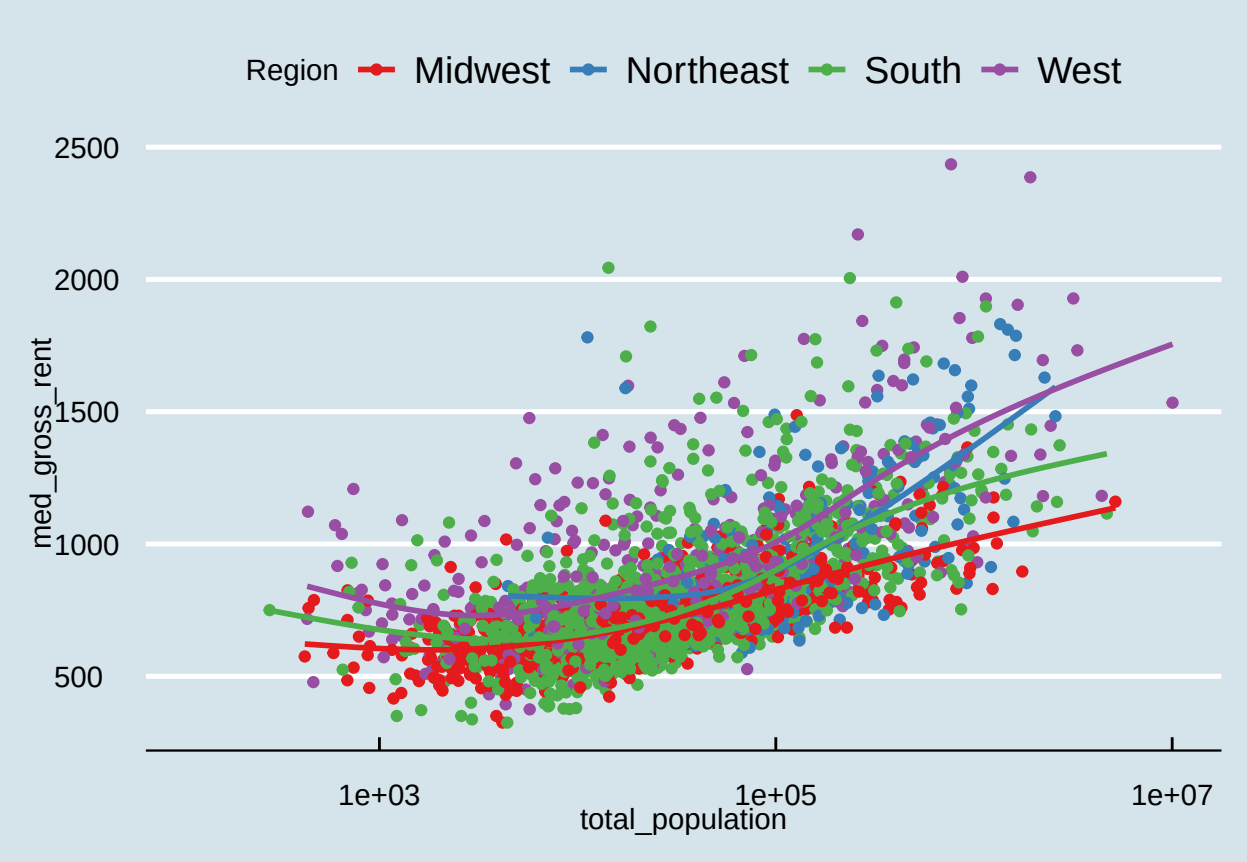
Question 2 - Basic ggplot

Using ggplot2, create a scatterplot of `med_incom` by `med_gross_rent`.

Question 3 - ggplot Reproduction

Using ggplot2, reproduce the scatterplot below. To help you do this, here are a few key notes:

- This visualizes `total_population` by `med_gross_rent`
- The dots are colored by the variable `Region`
- I am NOT showing the standard error of the LOESS smoother
- The X axis is displayed using a log scale
- I am using the "Set1" palette from Color Brewer
- I am using the Economist theme from the `ggthemes` package



Question 4 - Research Question

Oh no! Your try-hard problematic Gen X boss that wants to be a Gen Z just texted you:

> "Forgot to tell you about today's board meeting LOL! Need to know if there's any sort of relationship between race of a county and how they voted for pres in 2020 ASAP"

What do you send your boss? (note: don't overthink or overdo this. . . I just want to see you make a simple ggplot without any specific instructions)

BUTTERUP YOUR BOSS BONUS (not required): Go beyond what we discussed in the lecture and make this plot amazing. Potential additions: a title (+0.5 points), better labels (+0.5 points), adding color (+1 point), adding another variable using something other than color (+2 points).


ADDING A TITLE
BETTER LABELS
COLOR
FACETING
imgflip.com

```
# insert and run ggplot code here
```