

Lab 11 - Data, Aesthetics, & Geometries

Yana Chakalo

November 14, 2017

```
install.packages("readr")
```

```
## Installing package into '/Users/soniachakalo/Downloads/Final.Porject.Labs/packrat/lib/x86_64-apple-darwin18.0.0' (as 'lib' is unspecified)
```

```
##
```

```
## The downloaded binary packages are in
```

```
## /var/folders/c5/15rjsd_92hn_5rrsxxbmt_0r0000gn/T//Rtmpf1em75/downloaded_packages
```

```
install.packages("tidyverse")
```

```
## Installing package into '/Users/soniachakalo/Downloads/Final.Porject.Labs/packrat/lib/x86_64-apple-darwin18.0.0' (as 'lib' is unspecified)
```

```
##
```

```
## The downloaded binary packages are in
```

```
## /var/folders/c5/15rjsd_92hn_5rrsxxbmt_0r0000gn/T//Rtmpf1em75/downloaded_packages
```

Complete the following exercises below. Knit together the PDF document and commit both the Lab 11 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Which variables in your dataset are you interested in visualizing? Describe the level of measurement of these variables and what type of geography you think is appropriate to represent these variables. Give your reasoning for choosing the `geom_()` you selected.

The variables that I want to visualize are the frequency of the bias motives in each data set. For example, how many different types of bias motives are there in each year, so how many anti-black, anti-mental disability, anti-white, etc are there in each data set. I also want to look at the total frequency of what the offenders race is, so how many black offenders, how many white offenders, asian offenders, etc were there in each dataset. I also want to look at the frequency of city/states had the most incidents for each data set. And lastly, I want to look at the frequency of where these incidents occurred, like what were the highest and lowest places that these hate crimes took place. All of these variables are categorical so the best geography to represent these variables is a bar graph/histogram because this geom allows me to plot one continuous x variable and I will be able to see the frequency really well between different bias motives or offenders race, etc.

2. Is your data in the proper format to visualize the data in the way you want? Why or why not? *If you need/want to change the structure of your data, do it below.*

My data is in the format that i want to visualize all the variables I want because they are all categorical so I am able to look at the counts of each variable in each table thus giving me the frequency of the objects i want to see.

3. Create at least two different exploratory plots of the variables you chose using the skills we covered in class today. What types of mapping aesthetics did you choose and why? What do these plots tell you about your data?

The type of mapping aesthetic i chose was to put the Bias motive on the bottom, x axis, and then the frequency on the y axis. I wanted to choose as simple mapping as possible and not make it flashy or colorful. These two bar graphs tells us that there are a roughly similar bias motives committed in both years. There is one obvious bias motive that is the highest and then the other ones are much less. However, one big difference is that the bias motives for 2002 bar graph are more spread out and have more frequency among a wider

range of bias motives but for the 2013 data the bars are much less diverse and have a few key bias motives that are present the most showing that hate crimes have become much more targeted in 2013.

Making all the plots that i want for both data sets

```
install.packages("ggplot2")
```

```
## Installing package into '/Users/soniachakalo/Downloads/Final.Porject.Labs/packrat/lib/x86_64-apple-darwin18.0.0/lib' (as 'lib' is unspecified)
```

```
##
```

```
## The downloaded binary packages are in
```

```
## /var/folders/c5/l5rjsd_92hn_5rrsxxbmt_Or0000gn/T//Rtmpf1em75/downloaded_packages
```

```
data2 <- read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_23625_2002/DS0002/23625-0002-Data.tsv")
```

```
## Error in read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_23625_2002/DS0002/23625-0002-Data.tsv"): could not find function "read_tsv"
```

```
data13 <- read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_36118_2013/DS0002/36118-0002-Data.tsv")
```

```
## Error in read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_36118_2013/DS0002/36118-0002-Data.tsv"): could not find function "read_tsv"
```

Making my original data subset

```
data_subset2 <- data2 %>%
```

```
  select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, BIASMO1, VTYP_I1, LOCCOD1, VTYP_I1, LOCCOD1)
```

```
## Error in data2 %>% select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, : could not find function "%>%"
```

```
data_subset13 <- data13 %>%
```

```
  select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, BIASMO1, VTYP_I1, LOCCOD1, VTYP_I1, LOCCOD1)
```

```
## Error in data13 %>% select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, : could not find function "%>%"
```

2002 data set

```
ggplot(data = data_subset2,  
       mapping = aes(x = BIASMO1)) +  
  geom_bar()
```

```
## Error in ggplot(data = data_subset2, mapping = aes(x = BIASMO1)): could not find function "ggplot"
```

2013 data set

```
ggplot(data = data_subset13,  
       mapping = aes(x = BIASMO1)) +  
  geom_bar()
```

```
## Error in ggplot(data = data_subset13, mapping = aes(x = BIASMO1)): could not find function "ggplot"
```

4. Create at least three variations of the plots you've already made by modifying some of the arguments we covered in class (i.e. `position`, `scale`, `size`, `linetype` etc.). Do any of these modifications help you understand your data better? Why or why not? Do any of them create a misleading interpretation of the relationships between your variables? If yes, how so?

I tried doing modifications to my bar plots but they did not help create a more comprehensive plot. I think the scale is off and there is no definite separation from different types of bias motives.

2002

```
ggplot(data = data_subset2,
       mapping = aes(x = BIASM01)) +
  scale_x_continuous(
    name = "BIASM01",
    limits = c(NA),
    breaks = seq(0, 5, 10)) +
  geom_bar()
```

```
## Error in ggplot(data = data_subset2, mapping = aes(x = BIASM01)): could not find function "ggplot"
```

2013

```
ggplot(data = data_subset13,
       mapping = aes(x = BIASM01)) +
  scale_x_continuous(
    name = "BIASM01",
    limits = c(NA),
    breaks = seq(0, 5, 10)) +
  geom_bar()
```

```
## Error in ggplot(data = data_subset13, mapping = aes(x = BIASM01)): could not find function "ggplot"
```

5. From the plots you've created thus far, do any of them seem appropriate for a general audience? Why or why not? If so, what do you think you'd still need to do to make them more suitable as explanatory visualizations?

The only plots that I have made so far that might be appropriate for an audience are the first plots from question 3. They actually show a differentiation between bars and frequency's where you can see patterns and distinctions. I do however have to find a way to plot every type of bias motive to show individual frequencies which I can then compare to each other and to the other data set.

```
ggplot(data = data_subset13,
       mapping = aes(x = BIASM01, y = frequency(0, 1000))) +
  scale_x_continuous(
    name = "BIASM01",
    limits = c(NA),
    breaks = seq(0, 5, 10)) +
  geom_line()
```

```
## Error in ggplot(data = data_subset13, mapping = aes(x = BIASM01, y = frequency(0, : could not find f
```