



Chart Feedback

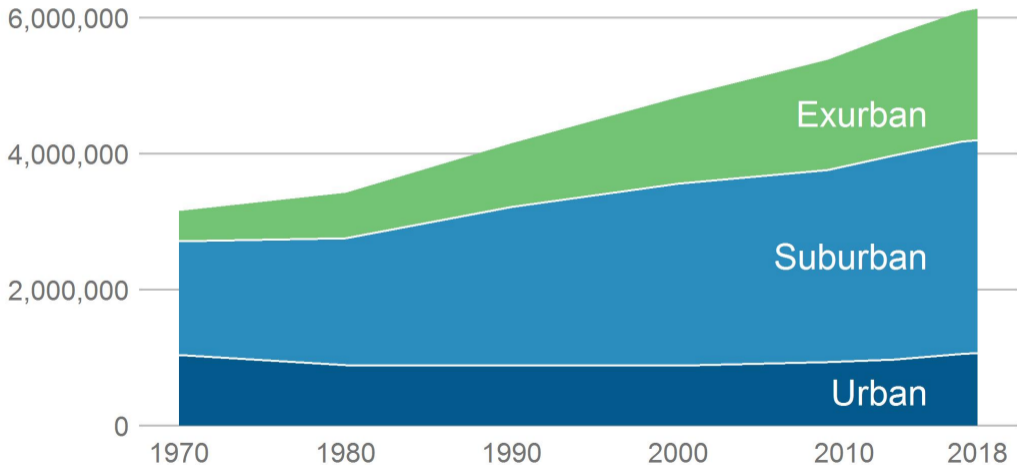
- Most of you could make a graph!
- With novel data!
- Some success
- But the last 10% takes 60 to 70% of the work
- No charts all the way there



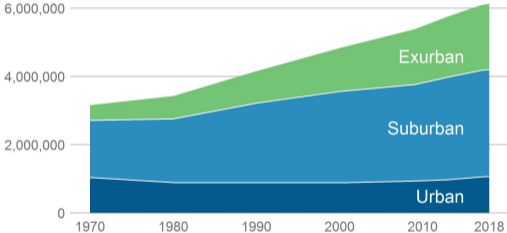
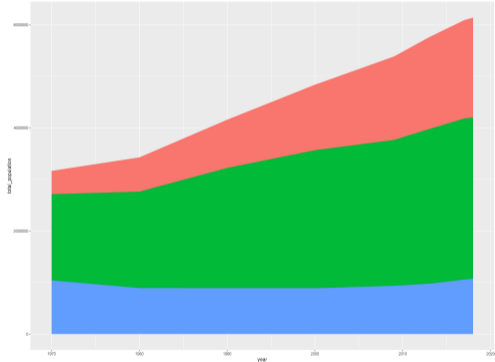
What is the Point?

“I’m not sure what it is you want readers to take away from this chart, so figuring out what that is, and then putting in the main title of the chart, is my first piece of advice.”

A Before and After: After



A Before and After: Both

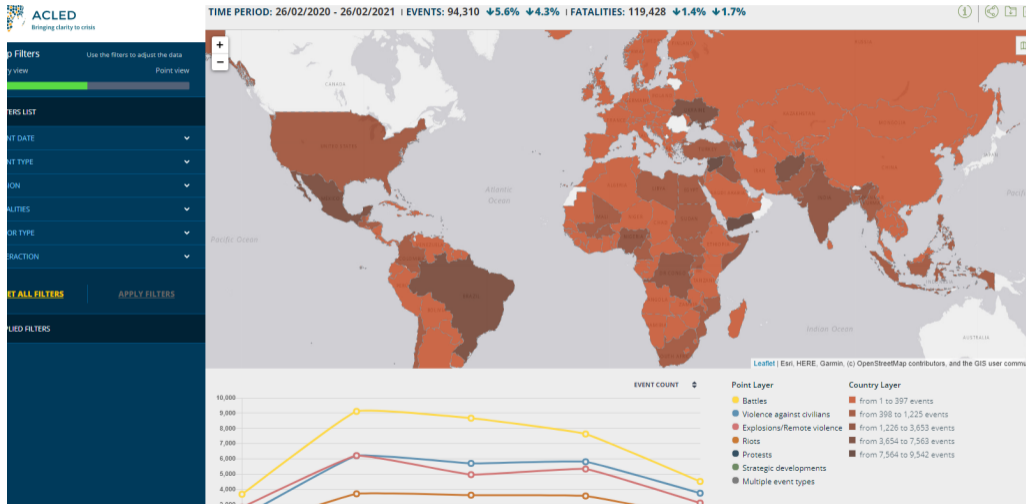




Helpful Standardizations

- Make a theme that you can use and modify for all of your charts
- Define a set of colors that you'll use throughout
- Define axis size labels you'll use throughout

Maureen's Example: ACLED



Go yourself to <https://acleddata.com/dashboard/>

B.1. Graduated Symbols

- Use symbol of graduated size to convey size or number
- Plot symbol at center of polygon
- Or at point location
- Used to convey absolute magnitudes – examples?
 - area
 - number of people
 - total home value

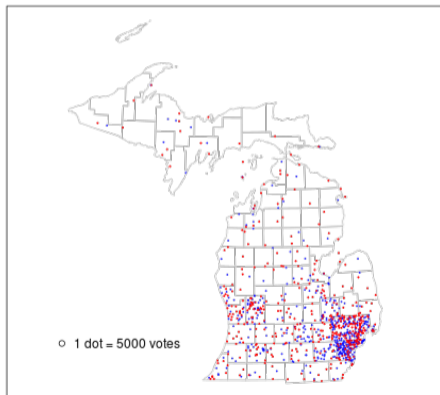
Strengths and Weaknesses of Graduated Symbol Maps

What do you think?

- Strengths
 - Disassociates area of administrative unit from magnitude conveyed
 - One of few methods for conveying absolute magnitude geographically
- Weaknesses
 - Can be hard to see all areas
 - 2-D size frequently not interpreted quantitatively appropriately

And With Two Variables

Michigan 2016 Election Dot Density Map



From https://msu.edu/~ashton/classes/866/notes/lect20/dot_mapping.html



Strengths and Weaknesses

Dot Density Best Practices

- Use only when geographical granularity of data approximate granularity of depiction
- Use color as in our upcoming discussion of choropleth maps

B.3. Choropleth Maps

- Used to show relative rates or intensities across space
- Examples?

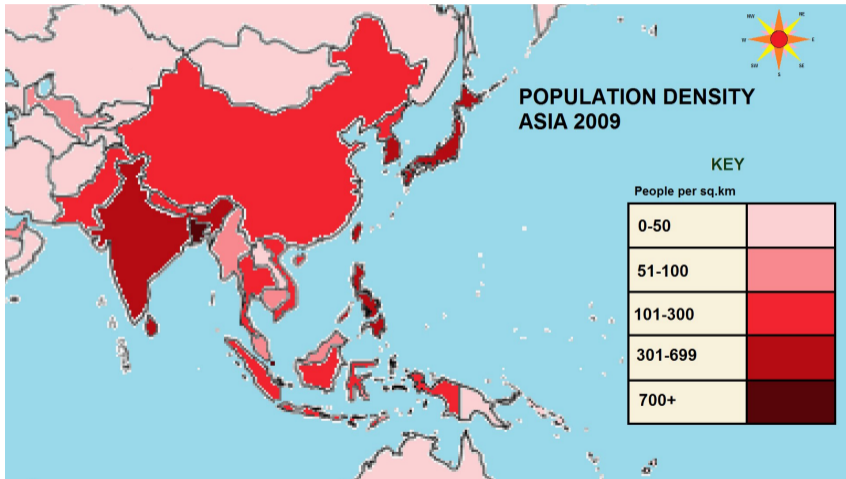
B.3. Choropleth Maps

- Used to show relative rates or intensities across space
- Examples?
 - population density
 - share in poverty
 - share covered by health insurance
- these can be continuous: unclassed
- or broken up into categories: classed
- Also used to show categorical differences across space
- Examples?

B.3. Choropleth Maps

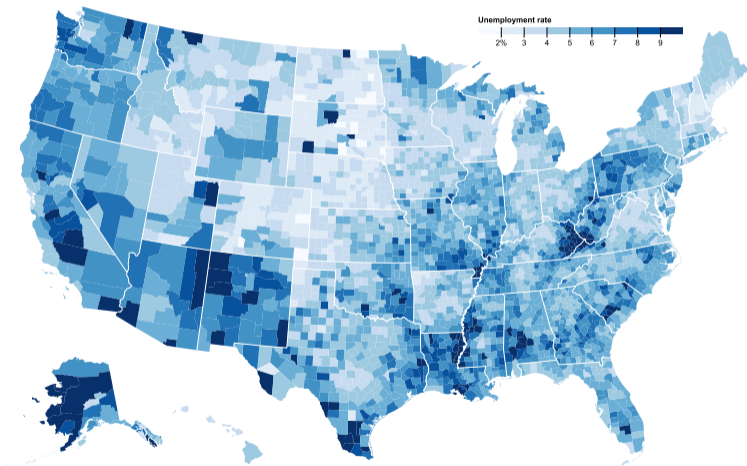
- Used to show relative rates or intensities across space
- Examples?
 - population density
 - share in poverty
 - share covered by health insurance
- these can be continuous: unclassed
- or broken up into categories: classed
- Also used to show categorical differences across space
- Examples?
 - ACA adoption or not
 - type of procurement legislation

Choropleth with Intensity



From <https://www.youtube.com/watch?v=PkmAiINPdrI>

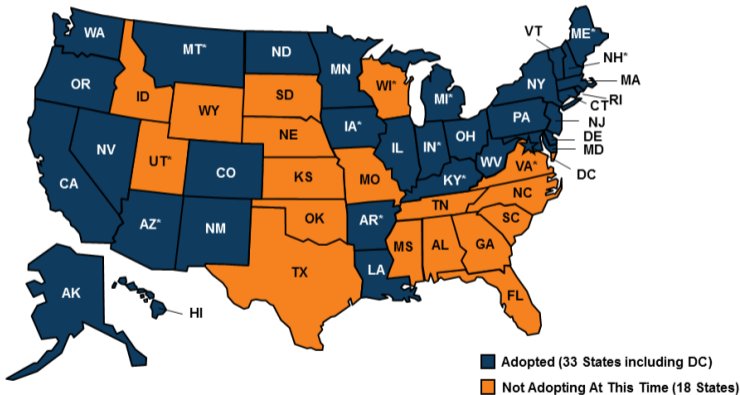
Choropleth with Intensity



From <https://bl.ocks.org/mbostock/4060606>

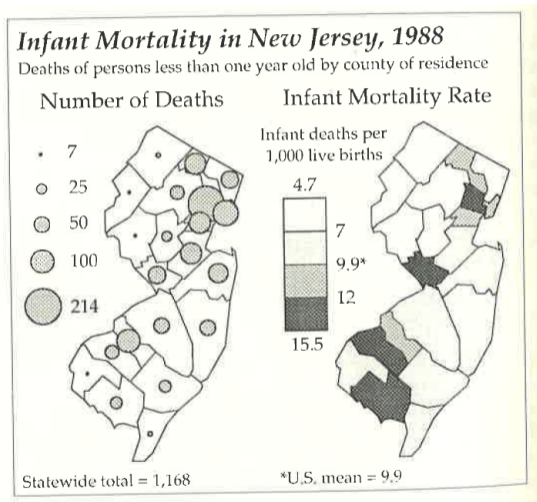
Categorical Map

Current Status of State Medicaid Expansion Decisions



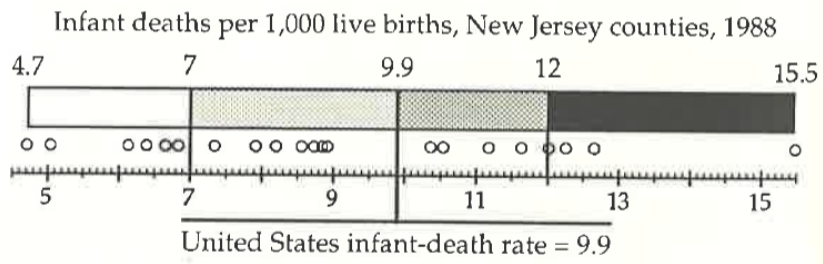
NOTES: Current status for each state is based on KFF tracking and analysis of state activity. *AR, AZ, IA, IN, KY, MI, MT, and NH have approved Section 1115 expansion waivers. VA is considering adopting expansion in their FY 2019 state budget and UT passed a law directing the state to seek CMS approval to partially expand Medicaid to 100% FPL using the ACA enhanced match; see the link below for more detail. CMS approved the Kentucky HEALTH expansion waiver on January 12, 2018; implementation will begin in April 2018. ME adopted the Medicaid expansion through a ballot initiative in November 2017; the ballot measure requires submission of a state plan amendment (SPA) within 90 days and implementation of expansion within 180 days of the measure's effective date; however, the governor failed to meet the SPA submission deadline (April 3). WI covers adults up to 100% FPL in Medicaid, but did not adopt the ACA expansion.

B.4. Combination of Count and Intensity Information



Better Yet, the Histogram Legend

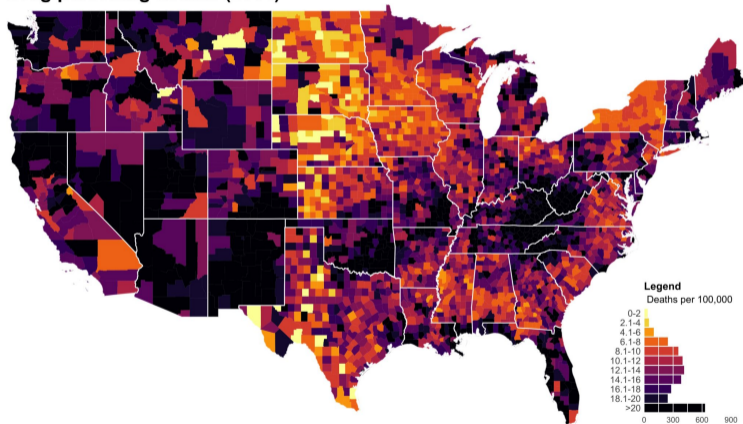
What does this add that the choropleth cannot convey?



Monomnier, Figure 6.10

Another Histogram Legend

Drug poisoning deaths (2014)



Source: <https://blogs.cdc.gov/hchs-data-visualization/drug-poisoning-mortality/>

Mapping Framework

C. Size vs Intensity

Monmonier on Count vs. Intensity Data

- Monmonier says never use a [what kind of map] for count data
 - Why?

Monmonier on Count vs. Intensity Data

- Monmonier says never use a [what kind of map] for count data
 - Why?
 - Because size should be the “principle visual variable” for such maps
- M. says use a choropleth for intensity

Monmonier on Count vs. Intensity Data

- Monmonier says never use a [what kind of map] for count data
 - Why?
 - Because size should be the “principle visual variable” for such maps
- M. says use a choropleth for intensity
- Agree with overall sentiment, but not sure it holds in all cases

Mapping Framework

D. Best Practices

D. Best Practices

1. Categories
2. Colors
3. Histogram legend
4. Worst practices

D.1. Categories

- 4 is great
- Don't use more than 5 or 6
- Use an intensity ramp only when you care very little about the exact values
- And comparison between values

D.2 Colors

- Make the most intense color the largest value
- Avoid pattern fills if at all possible
- Make your legend a dot plot or histogram with the same colors
- Put anything else on map in a light color

Use ColorBrewer

- Named after Cynthia Brewer
- <http://colorbrewer2.org/>
- You say
 - number of classes
 - sequential or divergent or qualitative
 - multi-or single hue
 - your preferred color
 - color-blind friendly?
 - screen or printer?
 - and more...
- and it gives you a color scheme!

D.4. Worst Practices

- Rainbow colors for classification
- Ones that are frequently bad ideas
 - Map total amounts
 - “geographic features that are continuous in nature”



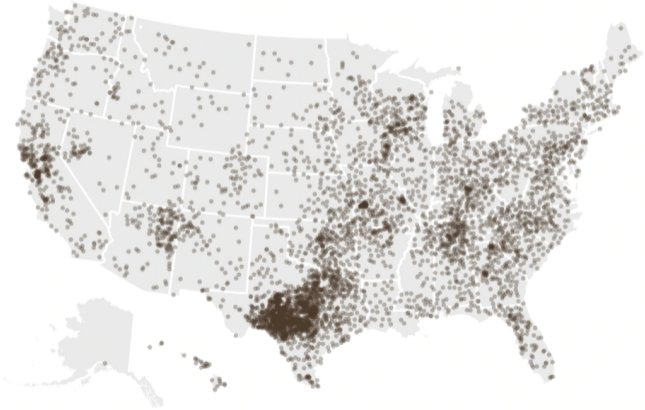
D.4. Worst Practices

- Rainbow colors for classification
- Ones that are frequently bad ideas
 - Map total amounts
 - “geographic features that are continuous in nature” “... because their distributions are not controlled by political or administrative subdivisions” (DTB, p. 104)

E. Think about Goats

“This is Literally Every Goat in the United States”

One dot = 500 goats.



WASHINGTONPOST.COM/**WONKBLOG**

Source: USDA Agricultural Census

Except it is not! See [WP article](#)



These Maps in R

Today in R Programming

- A. Choropleths
- B. Histogram legends
- C. Dot Density

A. Choropleth Maps

```
p1 <- ggplot() +  
  geom_sf(data = polys,  
          mapping = aes(fill = fill.in.variable))
```


Assigning Colors

Set colors at the beginning of your program:

```
my.4.colors <- c("c1", "c2", "c3", "c4")
```

Assigning Colors

Set colors at the beginning of your program:

```
my.4.colors <- c("c1", "c2", "c3", "c4")
```

Then use them wherever you'd like:

```
p1 <- ggplot() +  
  geom_sf(data = polys,  
          mapping = aes(fill = fill.in.variable)) +  
  scale_fill_manual(values = my.4.colors)
```

Assigning Colors

Set colors at the beginning of your program:

```
my.4.colors <- c("c1", "c2", "c3", "c4")
```

Then use them wherever you'd like:

```
p1 <- ggplot() +  
  geom_sf(data = polys,  
          mapping = aes(fill = fill.in.variable)) +  
  scale_fill_manual(values = my.4.colors)
```

Use the same set of colors for your map and your histogram legend

B. Histogram Legends

Why a histogram legend?

B. Histogram Legends

Why a histogram legend?

- ▶ Map gives a misleading or uninterpretable description of variable

B. Histogram Legends

Why a histogram legend?

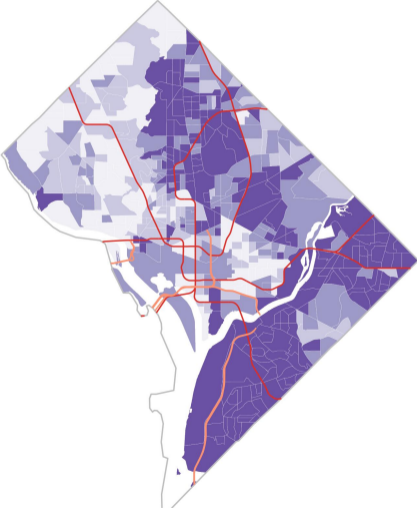
- ▶ Map gives a misleading or uninterpretable description of variable
- ▶ Can use a dot plot, as in Monmonier
 - ▶ `geom_dotplot()`
- ▶ Can use `geom_histogram()`
 - ▶ color bars with `fill`
- ▶ Can use `geom_bar()`

B. Histogram Legends

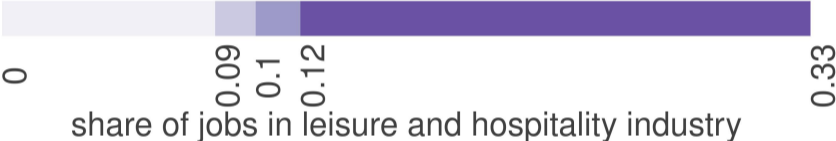
Why a histogram legend?

- ▶ Map gives a misleading or uninterpretable description of variable
- ▶ Can use a dot plot, as in Monmonier
 - ▶ `geom_dotplot()`
- ▶ Can use `geom_histogram()`
 - ▶ color bars with `fill`
- ▶ Can use `geom_bar()`
- ▶ **Colors must link with the plot – or the whole thing fails**

Example from DC: Can't Tell Magnitudes



Example from DC: Histogram Legend



Warning on the Histogram/Bar/Dot Legend

- ▶ need to trick R into making a bar chart
- ▶ assume we have four quartiles like this

```
ex.df[,c("quartile", "qmin", "qmax", "cato")]
```

```
##  quartile qmin qmax cato
## 1         1    0    1    1
## 2         2    1    2    1
## 3         3    2    4    1
## 4         4    4   23    1
```

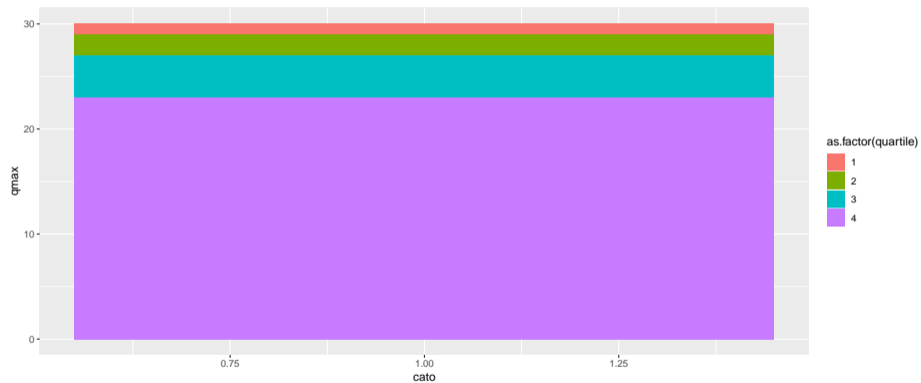
- ▶ plotting quartile vs max in a bar chart does not give what we want

The Bad Graph

```
p3 <- ggplot() +  
  geom_col(data = ex.df,  
           mapping = aes(x = cato, y = qmax,  
                         fill = as.factor(quartile)))
```

The Bad Graph

p3



We Need to Trick R

- ▶ add an “increment” variable
- ▶ and make the categories backwards

```
ex.df
```

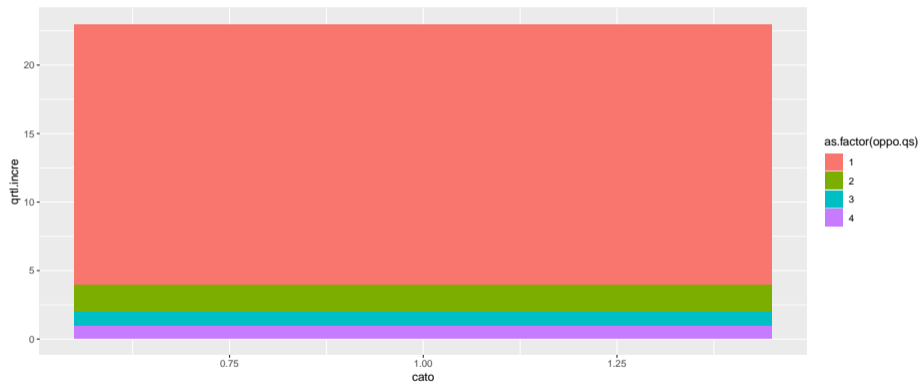
```
##   quartile cato qmin qmax oppo.qs qrtl.incre
## 1         1    1    0    1        4         1
## 2         2    1    1    2        3         1
## 3         3    1    2    4        2         2
## 4         4    1    4   23        1        19
```

The Better Graph

```
p4 <- ggplot() +  
  geom_col(data = ex.df,  
           mapping = aes(x = cato, y = qrtl.incre,  
                          fill = as.factor(oppo.qs)))
```

The Better Graph

p4

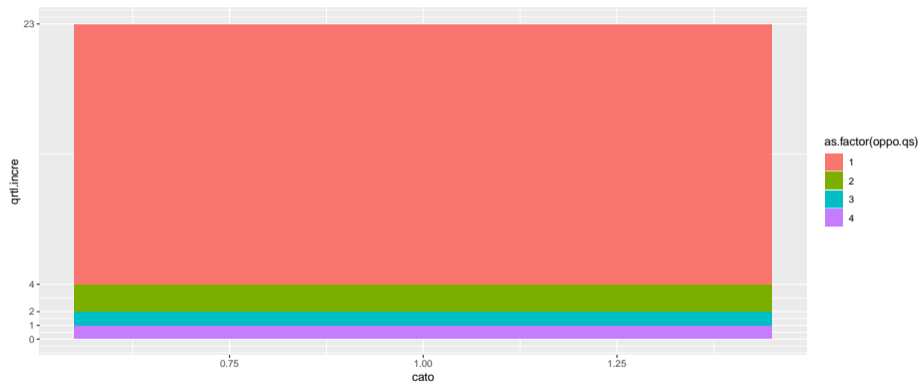


Further Trick R: Fix Labels

```
p4 <- ggplot() +  
  geom_col(data = ex.df,  
           mapping = aes(x = cato, y = qrtl.incre,  
                         fill = as.factor(oppo.qs))) +  
  scale_y_continuous(breaks = c(min(ex.df$qmin), ex.df$qmax),  
                    labels = c(min(ex.df$qmin), ex.df$qmax))
```


Further Trick R: Fix Labels

p4



C. Dot Density Maps

1. For each polygon, make a random sample of points based on scaling: `st_sample()`

C. Dot Density Maps

1. For each polygon, make a random sample of points based on scaling: `st_sample()`
2. Find the x/y coordinates of these points (in matrix form): `st_coordinates()`

C. Dot Density Maps

1. For each polygon, make a random sample of points based on scaling: `st_sample()`
2. Find the x/y coordinates of these points (in matrix form): `st_coordinates()`
3. Make the matrix a dataframe: `'as.data.frame()`

C. Dot Density Maps

1. For each polygon, make a random sample of points based on scaling: `st_sample()`
2. Find the x/y coordinates of these points (in matrix form): `st_coordinates()`
3. Make the matrix a dataframe: `'as.data.frame()'`
4. Plot the points with `ggplot`

C. Dot Density Maps

1. For each polygon, make a random sample of points based on scaling: `st_sample()`
2. Find the x/y coordinates of these points (in matrix form): `st_coordinates()`
3. Make the matrix a dataframe: `'as.data.frame()'`
4. Plot the points with `ggplot`
5. Make sure you plot your points on top of a map or they will look funny

Dot Density Map code

```
h.map <-  
  ggplot() +  
    # this is the map you did before  
    geom_sf(data = bg2010,  
            fill = "transparent",  
            color = "white") +  
    # here are the points  
    geom_point(data = hdat$df,  
              mapping = aes(x=lon, y = lat),  
              shape = ".")
```

