GEOG 491/891: Special Topics - Spatial Analysis in R

Week 02.02: A crash course in plotting data

Dr. Bitterman

Today's schedule

- Open discussion
- Plotting
- Exercises

Anything to discuss? Questions?

Why do we plot our data?

(I know it's a bit of a silly question, just go with it)

Many methods, we'll focus on two (but really on 1)

1. "base R"
2. ggplot

Setup

- Open RStudio, start a new project (or use the one from last class)
- Make sure you have `ne_counties.csv` from last class

Today's packages

```
library(tidyverse)
library(ggplot2)
```

read the data, remind yourself what it loiks like

```
mydf <- read_csv("./data/ne_counties.csv")
glimpse(mydf)
```

Some simple base R plots

the "plot" function can be used in multiple ways

```
# scatter plot
# should be highly correlated
plot(mydf$Total, mydf$TotalUnits)
```

What do you see?

```
# scatter plot #2
plot(mydf$Total, mydf$PerCapInc)
```

Let's try a histogram

```
# histogram
hist(mydf$PerCapInc)
```

and exert a bit more control

```
# change the number of breaks
hist(mydf$PerCapInc, breaks = 20)
```

ggplot

Let's build a plot step-by-step

the setup

```
# the initial call
ggplot(mydf, aes(x = Total, y = PerCapInc))
```

What happened?

Let's add a geom

Wait, what's a "geom"???

```
ggplot(mydf, aes(x = Total, y = PerCapInc)) +
  geom_point()
```

Note the "+" ...it's a very different notation

and we can modify the points:

```
ggplot(mydf, aes(x = Total, y = PerCapInc)) +
  geom_point(colour = "blue")
```

And alter the theme

```
ggplot(mydf, aes(x = Total, y = PerCapInc)) +
  geom_point(colour = "blue") +
  theme_minimal()
```

Try a different one! How would you know what options there are for themes?

Add some labels

```
ggplot(mydf, aes(x = Total, y = PerCapInc)) +
  geom_point(colour = "blue") +
  theme_minimal() +
  labs(x = "Total Population", y = "Per capita income")
```

...and add a title

```
# and give it a title
ggplot(mydf, aes(x = Total, y = PerCapInc)) +
  geom_point(colour = "blue") +
  theme_minimal() +
  labs(x = "Total Population", y = "Per capita income",
       title = "My first ggplot")
```

fit a line

don't just throw arbitrary models/fits on your data

```
ggplot(mydf, aes(x = Total, y = PerCapInc)) +
  geom_point(colour = "blue") +
  geom_smooth(method = "glm", colour = "red") +
  theme_minimal() +
  labs(x = "Total Population", y = "Per capita income",
       title = "My first ggplot")
```

Let's try something different

Using categorical data

(first, we have to make some categories) - let's walk through this code

```
mydf2 <- mydf %>% mutate(sizeCategory = ifelse(Total > 20000, "big", "small"))
```

check your work:

```
summary(mydf2$sizeCategory) ### What happened?

# turn them into factors to count them
summary(as.factor(mydf2$sizeCategory))
```

A first example

Let's break it down:

```
ggplot(mydf2, aes(x = Total, y = PerCapInc)) +
  geom_point(aes(shape = sizeCategory, colour = sizeCategory), size = 3) +
  theme_minimal() +
  labs(x = "Total Population", y = "Per capita income",
       title = "My formatted ggplot")
```

what happened?

One more example:

a pipe with a boxplot

```
mydf2 %>% ggplot(., aes(x = sizeCategory, y = PerCapInc)) +
  geom_boxplot(aes(fill = sizeCategory)) +
  theme_minimal() +
  labs(x = "Categorical size",
       y = "Per capita income",
       title = "I made a boxplot",
       subtitle = "it's handy")
```

If there's time, try it yourself:

using ggplot:

- Make a histogram
- Make a histogram of Females with advanced degrees (multi-step process)
- Try a barplot...
    - subset of the counties (however you want)
    - counties on the x-axis, number of vacant uints on the y-axis

Review and next class

- Any questions?
- This week’s readings/tasks:
    - Chapter 2 in textbook
    - Continue to review Hadley's book/site
    - Practice on your own