COMPSCIX 415.2 Homework 3

Rajat Jain

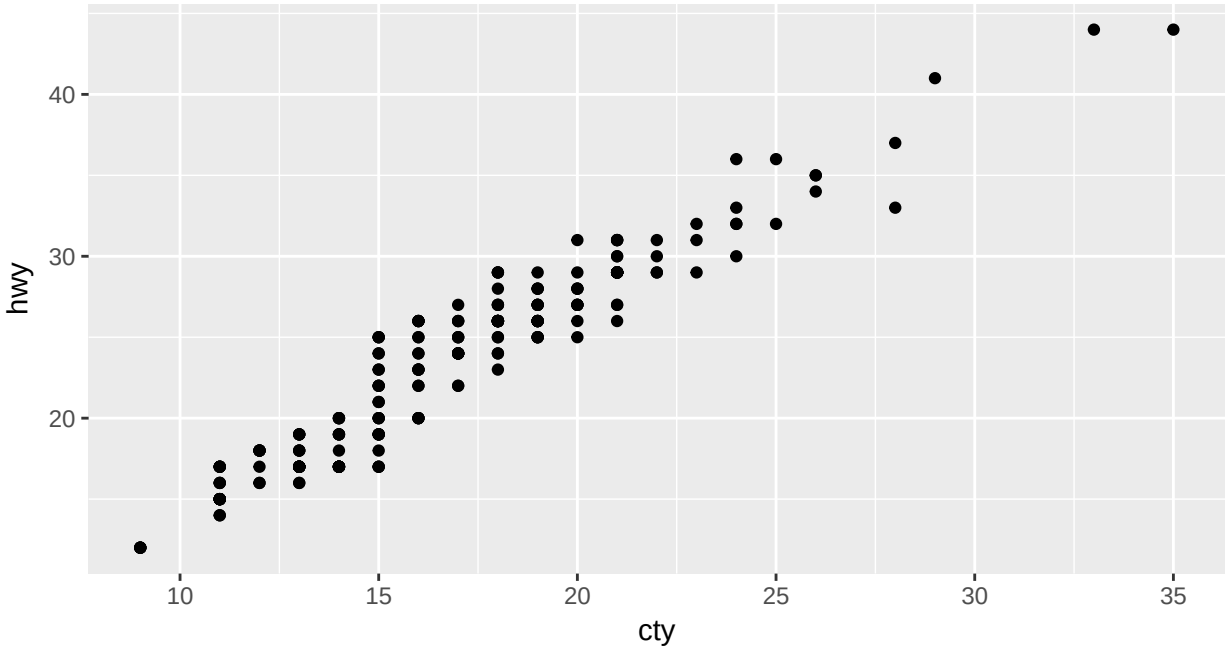
June 19th, 2018

Contents

Section 3.8.1 Exercises

1. **What is the problem with this plot? How could you improve it?**

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +
  geom_point()
```



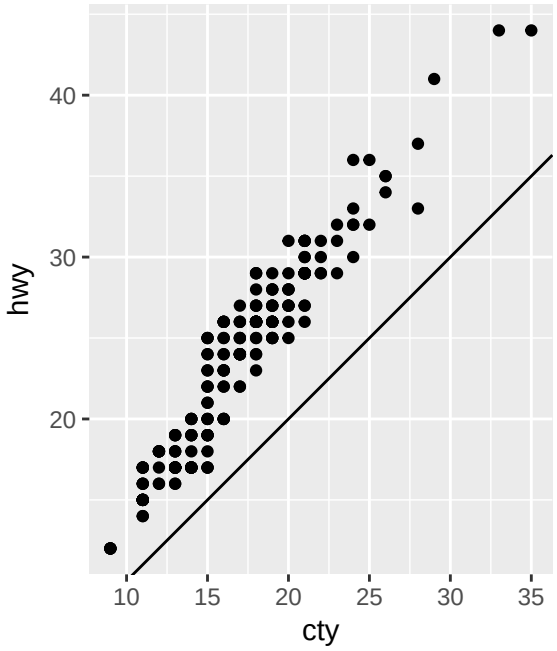
2. **What parameters to geom_jitter() control the amount of jittering?**
3. **Compare and contrast geom_jitter() with geom_count().**
4. **What's the default position adjustment for geom_boxplot()? Create a visualisation of the mpg dataset that demonstrates it.**

Section 3.9.1 Exercises

2. **What does labs() do? Read the documentation.**

3. **What does the plot below tell you about the relationship between city and highway mpg? Why is coord_fixed() important? What does geom_abline() do?**

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +
  geom_point() +
  geom_abline() +
  coord_fixed()
```



Section 4.4 Exercises

1. **Why does this code not work?**

```
my_variable <- 10
my_varıable
#> Error in eval(expr, envir, enclos): object 'my_varıable' not found```
```

Look carefully! (This may seem like an exercise in pointlessness, but training your brain to notice even the tiniest difference will pay off when programming.)

2. **Tweak each of the following R commands so that they run correctly:**

```
library(tidyverse)

ggplot(dota = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))

fliter(mpg, cyl = 8)
filter(diamond, carat > 3)
```

Section 5.2.4 Exercises

1. **Find all flights that**
    1. Had an arrival delay of two or more hours

2. Flew to Houston (IAH or HOU)
3. Were operated by United, American, or Delta
4. Departed in summer (July, August, and September)
5. Arrived more than two hours late, but didn't leave late
6. Were delayed by at least an hour, but made up over 30 minutes in flight
7. Departed between midnight and 6am (inclusive)

2. **How many flights have a missing dep_time? What other variables are missing? What might these rows represent?**
3. **Why is NA ^ 0 not missing? Why is NA | TRUE not missing? Why is FALSE & NA not missing? Can you figure out the general rule? (NA * 0 is a tricky counterexample!)**

Section 5.4.1 Exercises

1. **Brainstorm as many ways as possible to select `dep_time`, `dep_delay`, `arr_time`, and `arr_delay` from `flights`.**
2. **What does the one_of() function do? Why might it be helpful in conjunction with this vector?**

```
vars <- c("year", "month", "day", "dep_delay", "arr_delay")
```