

# Programming your Pictures: R

Aidan Delaney & Brent Yorgey

August 4, 2016

## Introduction

This is a worksheet to support the “Programming your Pictures” workshop at Diagrams 2016. There are two versions of this worksheet, one with answers and one without. If you’d like the version with the answers then have a look at <sup>1</sup>. In these questions we will use the mpg dataset shipped with ggplot2. You might also find the cheat sheet at <sup>2</sup> to be helpful.

## First Plot

### Exercise 1: Plot and colour

Plot the mpg of cars for all manufacturers. An example plot can be seen in figure 1. Can you add in some colour based on the number of cylinders in the individual model?

Furthermore, can the visualisation be coaxed to provide the specific model for the car with the highest mpg? The following will plot car mpg over the manufacturer

```
ggplot(mpg, aes(x=factor(manufacturer), y=hwy)) +  
  geom_point(aes(colour=factor(cyl))) + scale_  
  color_brewer(type="div")
```

If you want to unveil the name of the best performing model we can change our geom\_point to geom\_text. It’s not very pretty, but it works for data exploration.

```
ggplot(mpg, aes(x=factor(manufacturer), y=hwy)) +  
  geom_text(aes(label=model, colour=factor(cyl))) +  
  scale_color_brewer(type="div")
```

### Exercise 2: Histogram

Generate a histogram of the mpg values for cars. You are advised to examine the tutorial notes for an example and some help<sup>3</sup>.

The following will plot the histogram of all mpg values

```
ggplot(mpg, aes(x=hwy)) + geom_histogram()
```

It could be more useful if we consider cars to the nearest 5 mpg.

<sup>1</sup> <http://aidandelaney.github.io/handouts/2016DiagramsRTutorial-answers.pdf>

<sup>2</sup> <https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

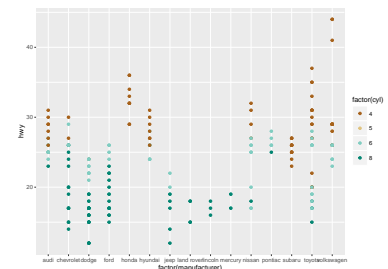


Figure 1: Plotting mpg for each manufacturer

<sup>3</sup> <http://aidandelaney.github.io/>

```
ggplot(mpg, aes(x=hwy)) + geom_histogram(binwidth
=5)
```

Or maybe you'd prefer minor breaks every 2 mpg?

```
ggplot(mpg, aes(x=hwy)) + geom_histogram(binwidth
=5) + scale_x_continuous(minor_breaks = c(2:max(
mpg$hwy)))
```

### *Exercise 3: Fit a linear model*

It might be useful to see the relationship between the cars' city and hwy mpg listings. Please generate a point plot of one over the other and fit a linear model to the points.

For this we have to plot both the points and the smoothed conditional mean. We also have to set the right method for what we want to achieve.

```
ggplot(mpg, aes(x=hwy, y=cty)) + geom_point() +
  geom_smooth(method="lm")
```

This plot might look better in a paper if there were less around it, and the line was dashed.

```
ggplot(mpg, aes(x=hwy, y=cty)) + geom_smooth(method
="lm", linetype="dashed") + theme_minimal()
```