

Problem Set - 1

Aaron An

September 15, 2016

Directions

During ANLY 512 we will be studying the theory and practice of *data visualization*. We will be using R and the packages within R to assemble data and construct many different types of visualizations. We begin by studying some of the theoretical aspects of visualization. To do that we must appreciate the actual steps in the process of making a visualization.

Most of us use software to do this and have done so for so long that we have lost an appreciation for the mechanistic steps involved in accurately graphing data. We will fix that this week by creating a series of analog (**meaning you draw them by hand**) graphics. The visualizations you create must be numerically and visually accurate and precisely scaled. Because of that the data sets we visualize will be small.

- A couple of tips, remember that there is preprocessing involved in many graphics so you may have to do summaries or calculations to prepare, those should be included in your work.
- To ensure accuracy you should draft your graphic and their supporting calculations using graph paper.

The final product of your home work (this file) should include scanned or photographed images for each question below and be submitted via Moodle. It is important that you name your file like this: YourName-ANLY512-50-FALL-2016. File naming is crucial and failing to do so results in a deduction.

I plotted them by hands and then used `ggplot` to check if I did it correctly, so might as well documented them here too.

Questions

1. Find the `mtcars` data in R. This is the dataset that you will use to create your graphics. Show a summary of the dataset.

```
mtcars %>% summary()
```

```
##      mpg          cyl          disp          hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
## 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
## Median :19.20   Median :6.000   Median :196.3   Median :123.0
## Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
## 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
## Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##      drat          wt          qsec          vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
## 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
## Median :3.695   Median :3.325   Median :17.71   Median :0.0000
## Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
## 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
## Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##      am          gear          carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
## 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
## Median :0.0000   Median :4.000   Median :2.000
```

```
## Mean :0.4062 Mean :3.688 Mean :2.812
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :1.0000 Max. :5.000 Max. :8.000
```

2. Draw a bar graph, that shows the number of each gear type in mtcars.

```
mtcars %>%
  ggplot(aes(factor(gear))) +
  geom_bar(width = 0.4, fill = "#004D43", alpha = 0.8)
```

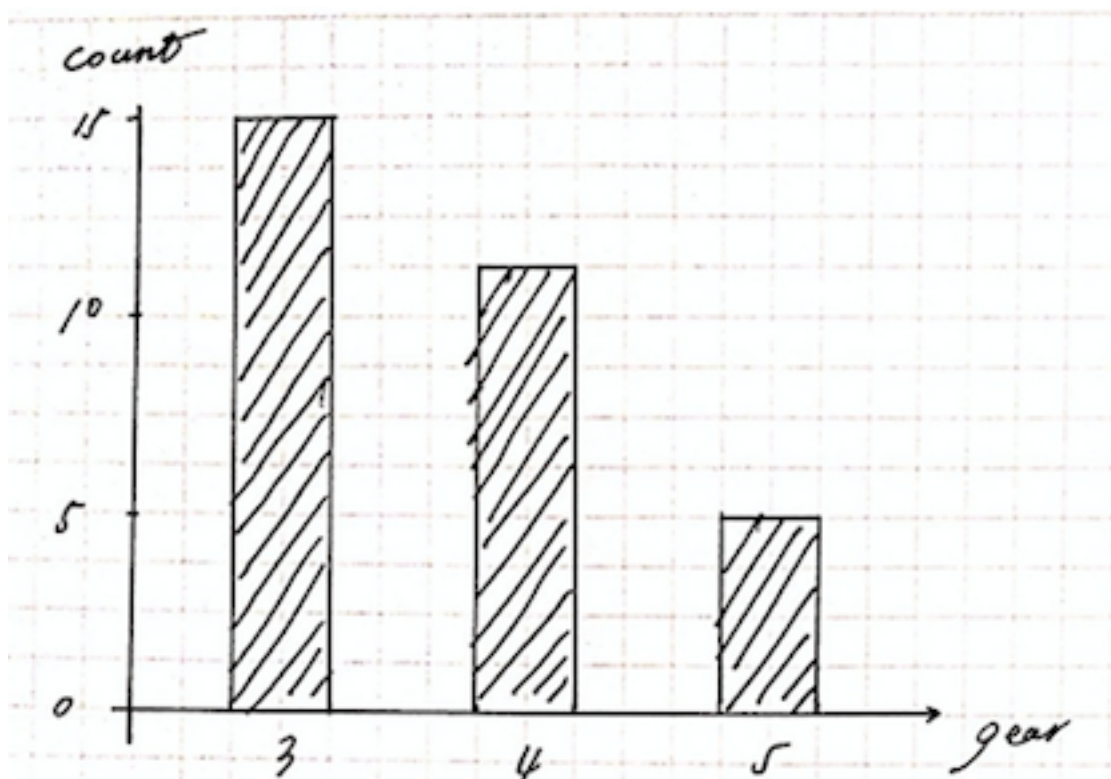
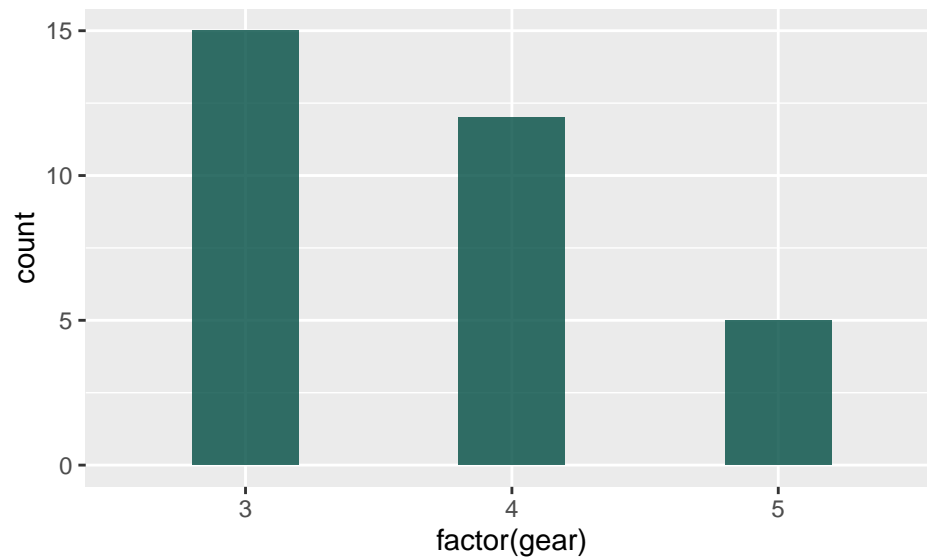


Figure 1:

- Next show a stacked bar graph of the number of each `gear` type and how they are further divided out by `cyl`.

```
mtcars %>%
  ggplot(aes(factor(gear), fill = factor(cyl))) +
  geom_bar(width = 0.4) + scale_fill_brewer(palette = "Greens")
```

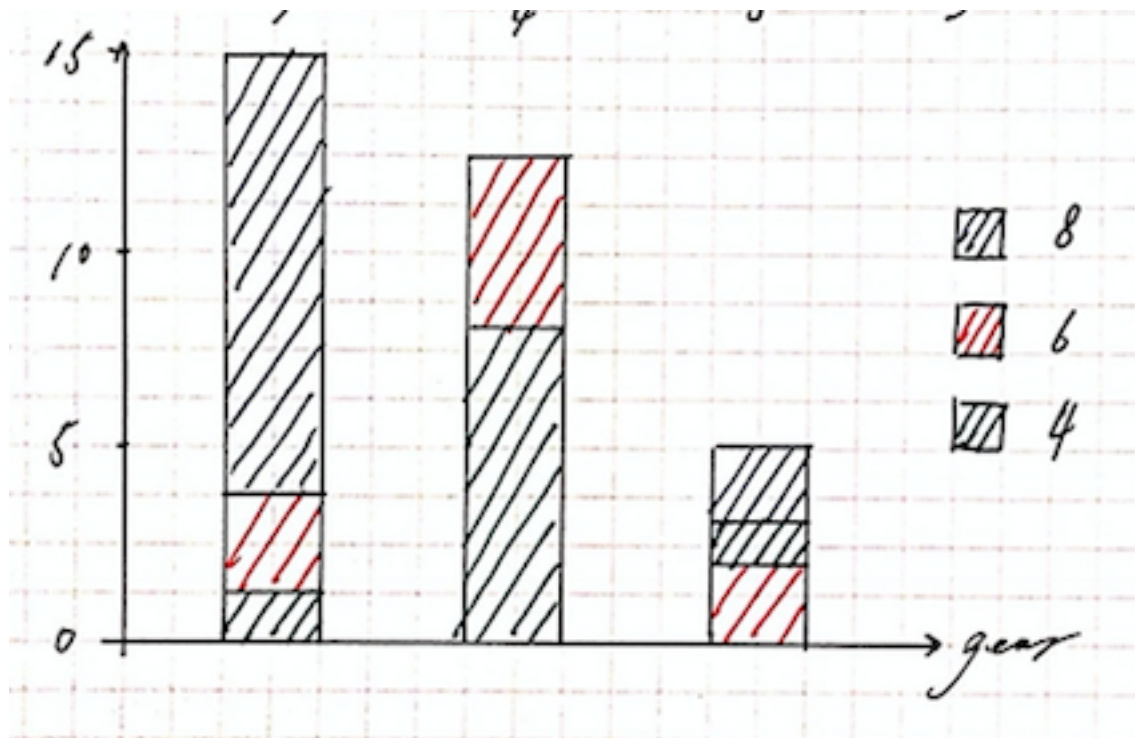
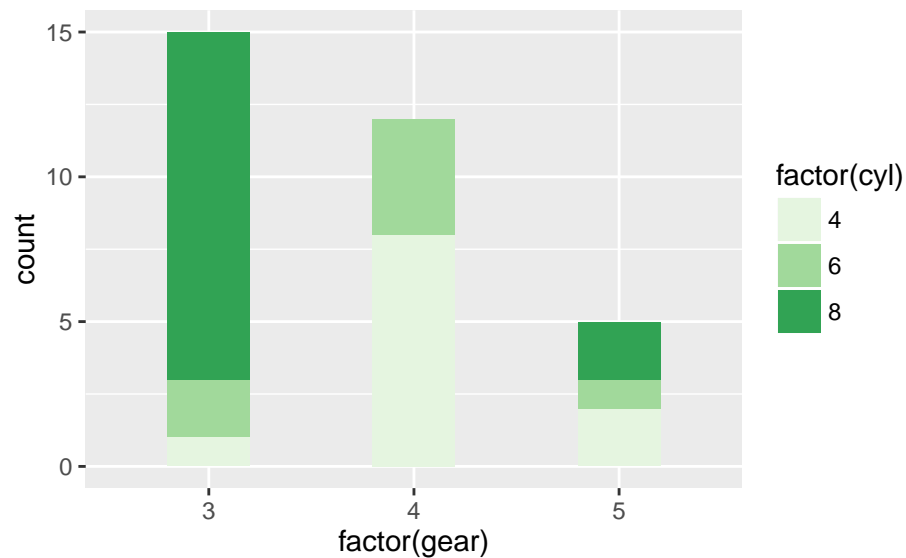


Figure 2:

- Draw a scatter plot showing the relationship between `wt` and `mpg`.

```
mtcars %>%
  ggplot(aes(wt, mpg)) +
```

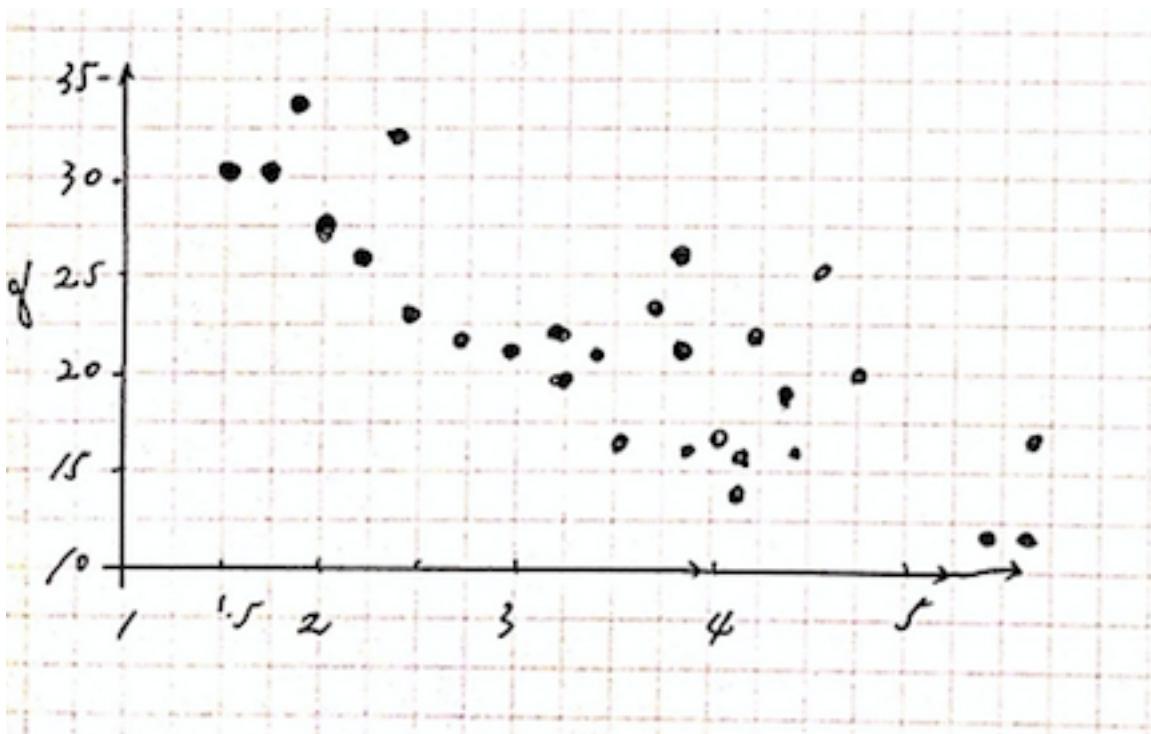
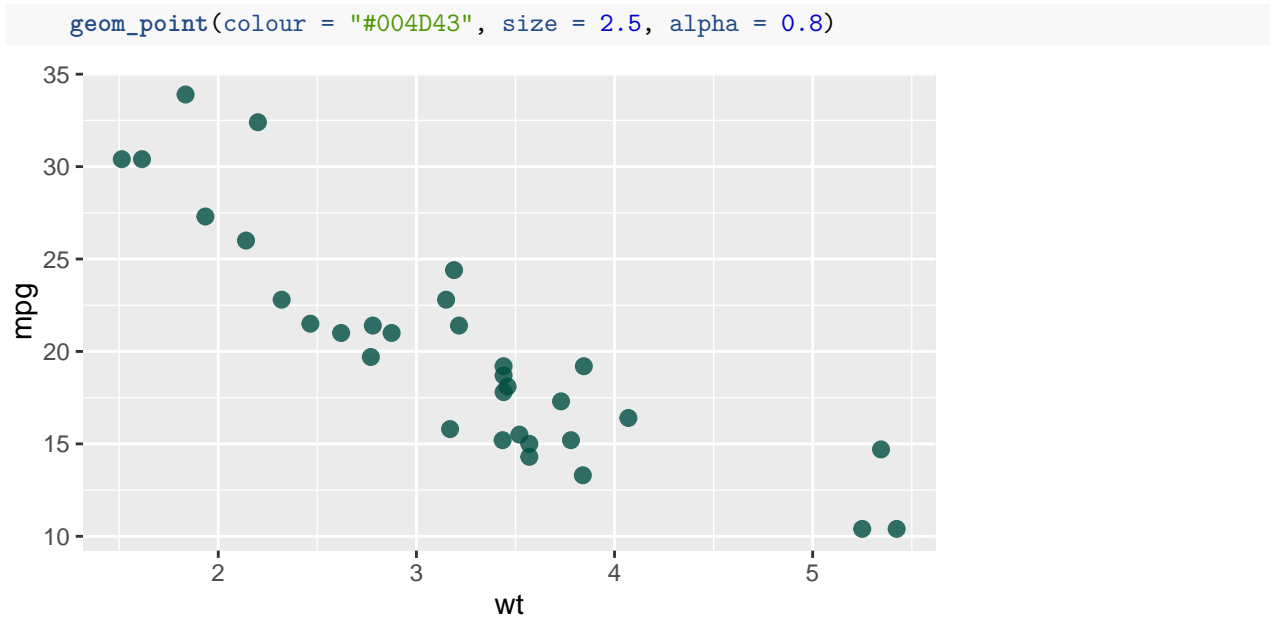


Figure 3:

- Finally draw a pie chart showing the proportion of cars from the `mtcars` data set that have different `carb` values.

```
mtcars %>%
  ggplot(aes(x = factor(1), fill = factor(carb))) +
  geom_bar(width = 1) +
  coord_polar(theta = "y") +
  scale_fill_brewer(palette = "Greens")
```

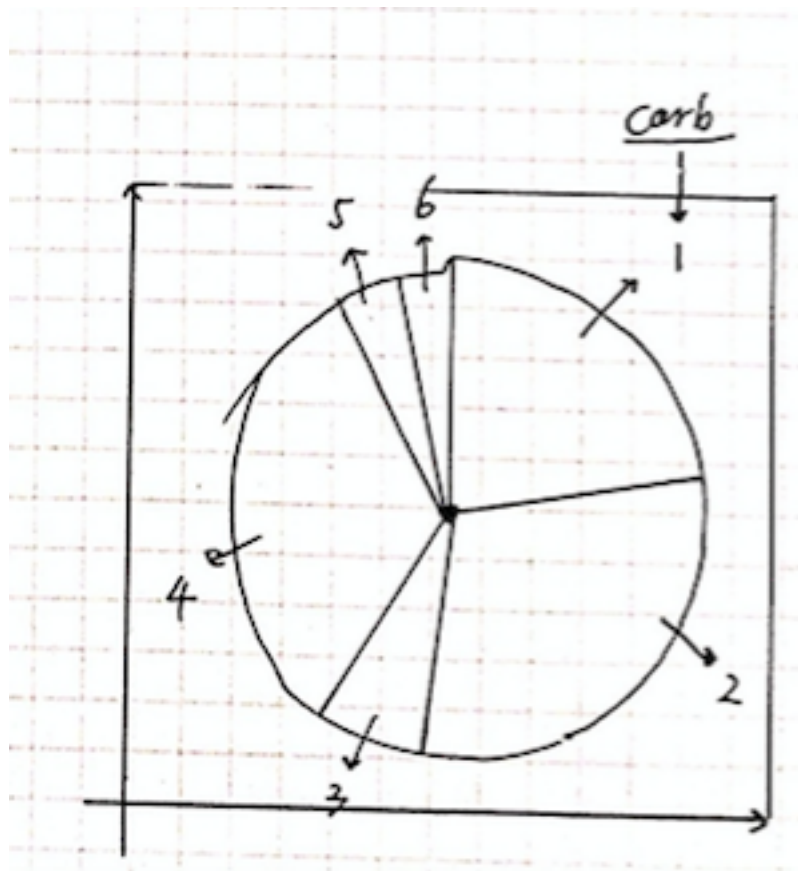
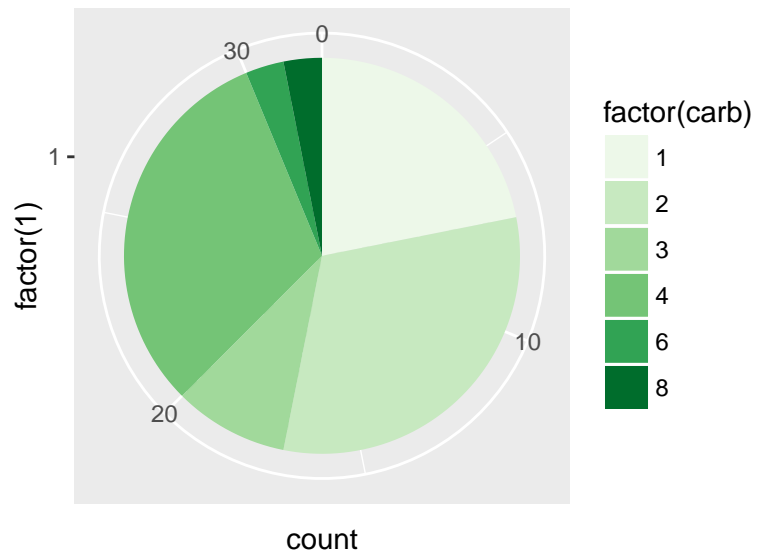


Figure 4: