MATH 118: Notes F

Making plots with ggplot2: line graphs, histograms & boxplots

This dataset contains a subset of the fuel economy data that the EPA makes available on https://fueleconomy.gov/. It contains only models which had a new release every year between 1999 and 2008 - this was used as a proxy for the popularity of the car.

```
library(tidyverse)
data("mpg")
```

Line Graphs

```
table <- mpg %>%
  group_by(cyl, class) %>%
  summarize(mean_hwy = mean(hwy))

table
```

```
## # A tibble: 19 x 3
## # Groups:   cyl [4]
##      cyl class      mean_hwy
##    <int> <chr>         <dbl>
##  1     4 compact        29.5
##  2     4 midsize        29.2
##  3     4 minivan        24
##  4     4 pickup         20.7
##  5     4 subcompact     30.8
##  6     4 suv            23.8
##  7     5 compact        29
##  8     5 subcompact     28.5
##  9     6 compact        25.3
## 10     6 midsize        26.3
## 11     6 minivan        22.2
## 12     6 pickup         17.9
## 13     6 subcompact     24.7
## 14     6 suv            18.5
## 15     8 2seater        24.8
## 16     8 midsize        24
## 17     8 pickup         15.8
## 18     8 subcompact     21.6
## 19     8 suv            16.8
```
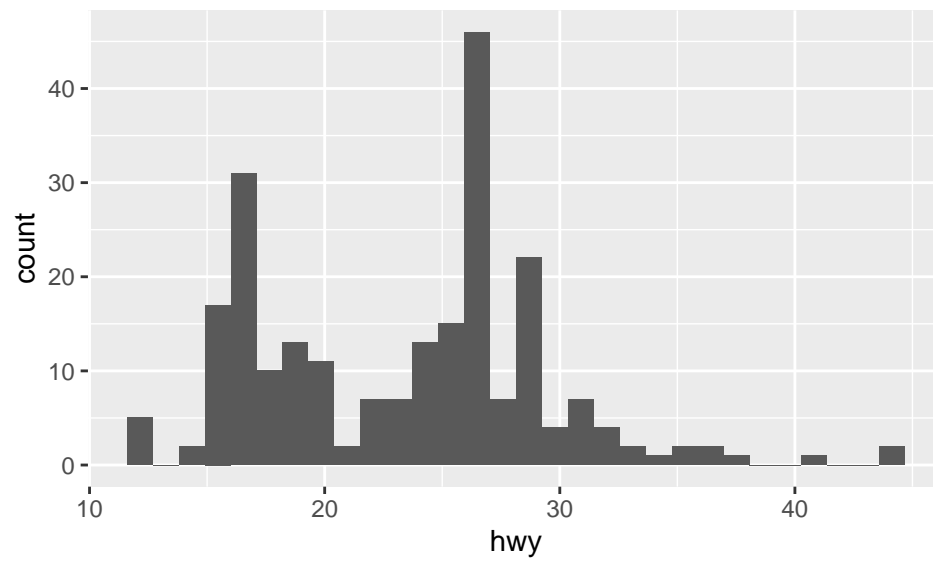
```
table %>%
  ggplot(aes(x=cyl, y=mean_hwy, group=class, color=class)) +
  geom_line()
```
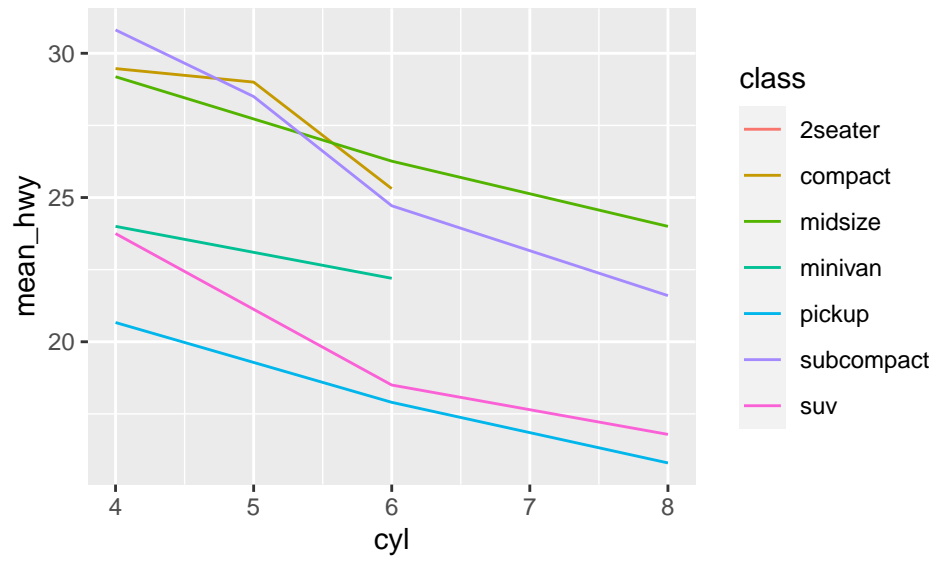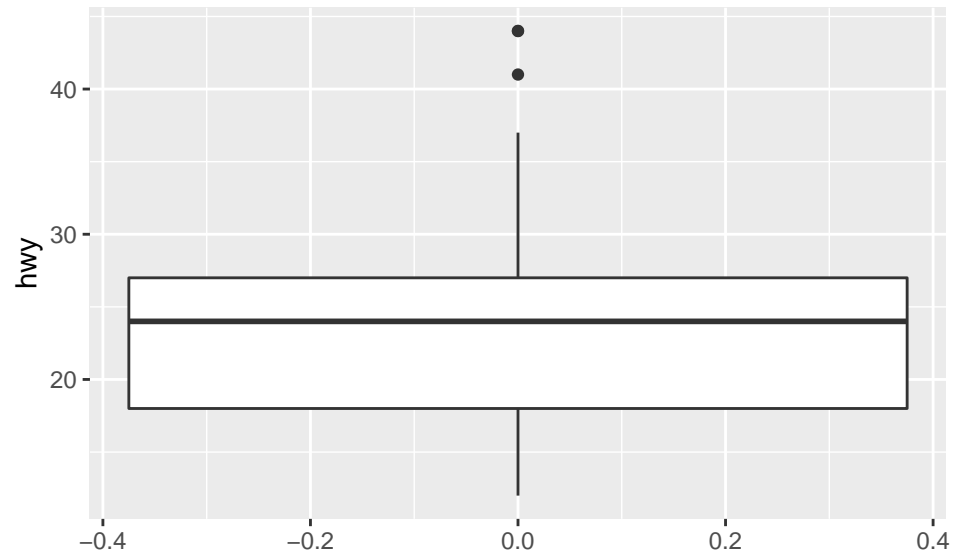
Histograms

```
mpg %>%
  ggplot(aes(x=hwy)) +
  geom_histogram()
```
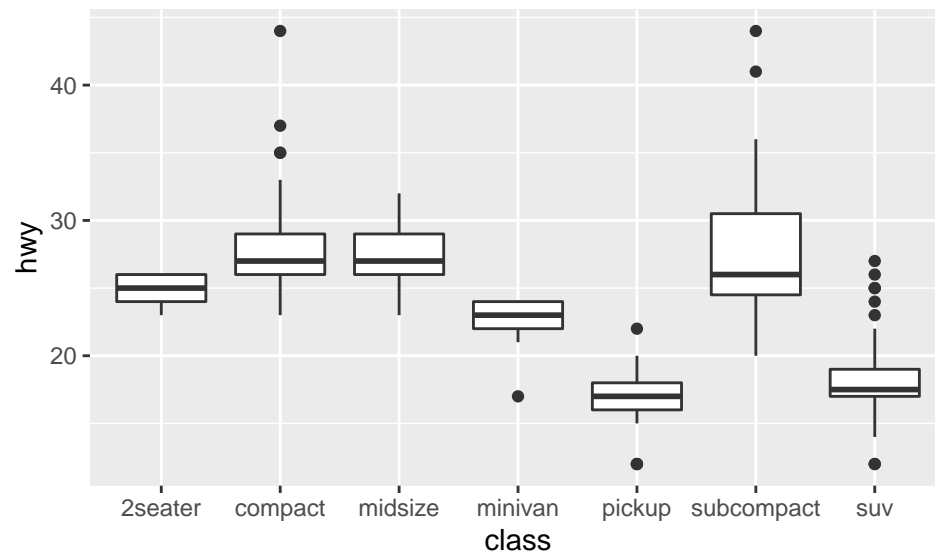
Boxplots

```
mpg %>%
  ggplot(aes(y=hwy)) +
  geom_boxplot()
```



```
mpg %>%
  ggplot(aes(x=class, y=hwy)) +
  geom_boxplot()
```
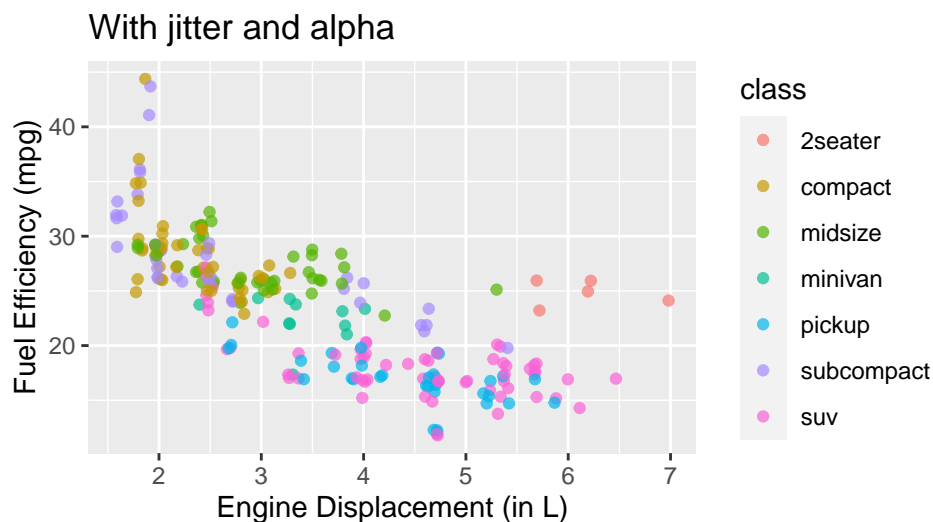
Jitter

Jittering is a technique for adding random noise to data points that have identical values in a plot. It is a useful method for avoiding overplotting and making it easier to visualize the density of data points that are otherwise hidden behind each other.

```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  ggtitle("Overlapping Points")
```
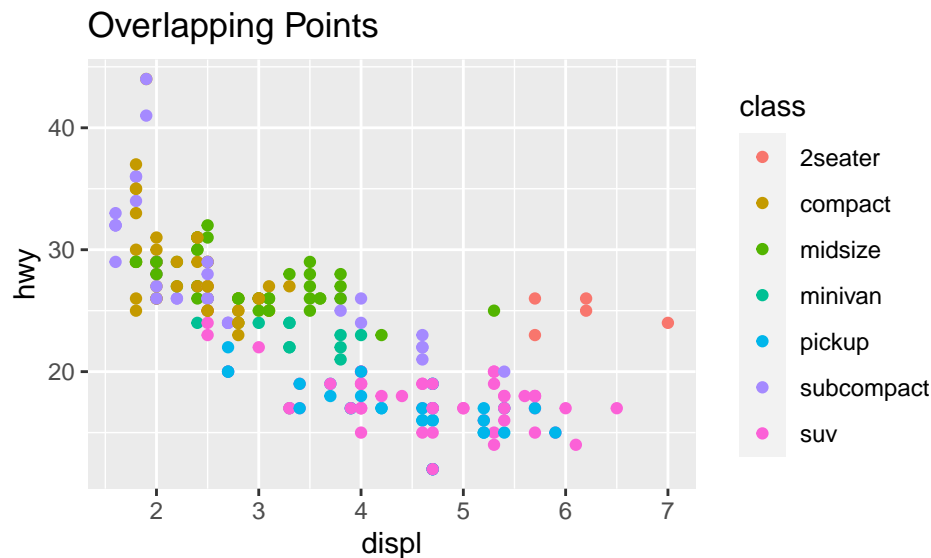


```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class), position = "jitter", alpha=0.7) +
  ggtitle("With jitter and alpha")  +
  xlab("Engine Displacement (in L)") +
  ylab("Fuel Efficiency (mpg)") +
  labs(caption = "Source: Fuel Economy Data 1999 - 2008 www.fueleconomy.gov")
```
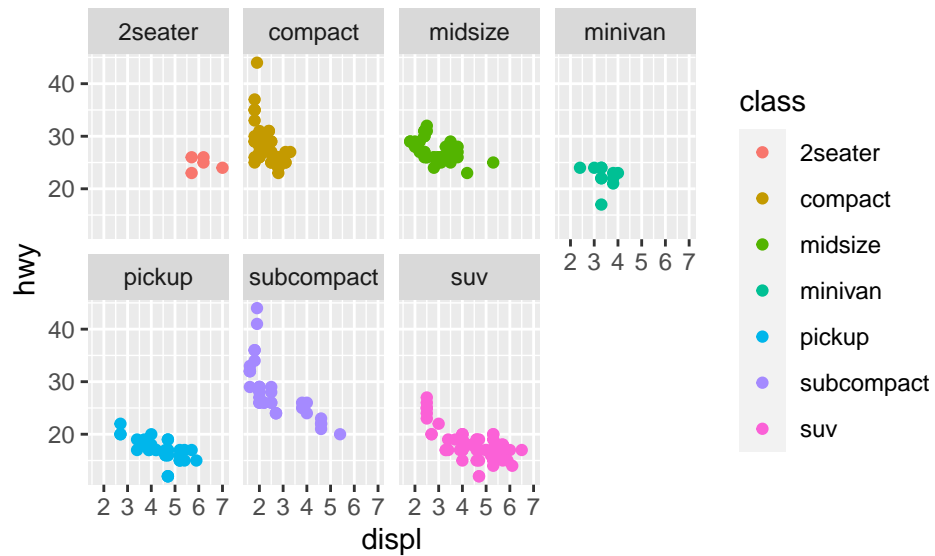
Facet Wrap

facet_wrap() is a function in the `ggplot2` package that allows you to create a multi-panel plot showing a similar plot over different subsets of the data, usually different values of a categorical variable.

```
mpg %>%
  ggplot() +
    geom_point(aes(x = displ, y = hwy, color=class)) +
    facet_wrap(~ class, nrow = 2)
```
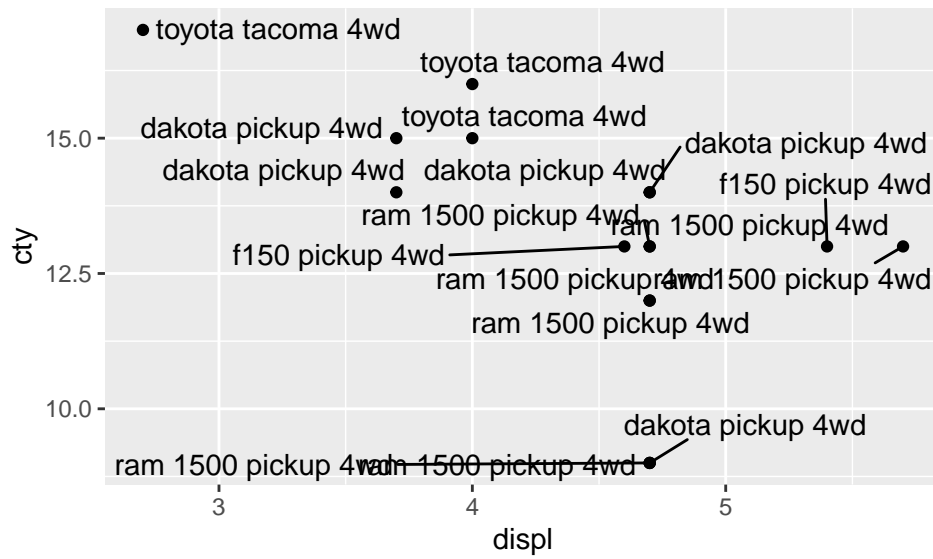
Labels

Adding text to a plot is one of the most common forms of annotation. Most plots will not benefit from adding text to every single observation on the plot, but labelling outliers and other important points is very useful.

```
mpg %>%
  filter(class=="pickup") %>%
  filter(year == 2008) %>%
  ggplot(aes(x=displ, y=cty)) +
    geom_point() +
    geom_text(aes(label=model),
              nudge_x = 0.25,
              nudge_y=0.25)
```
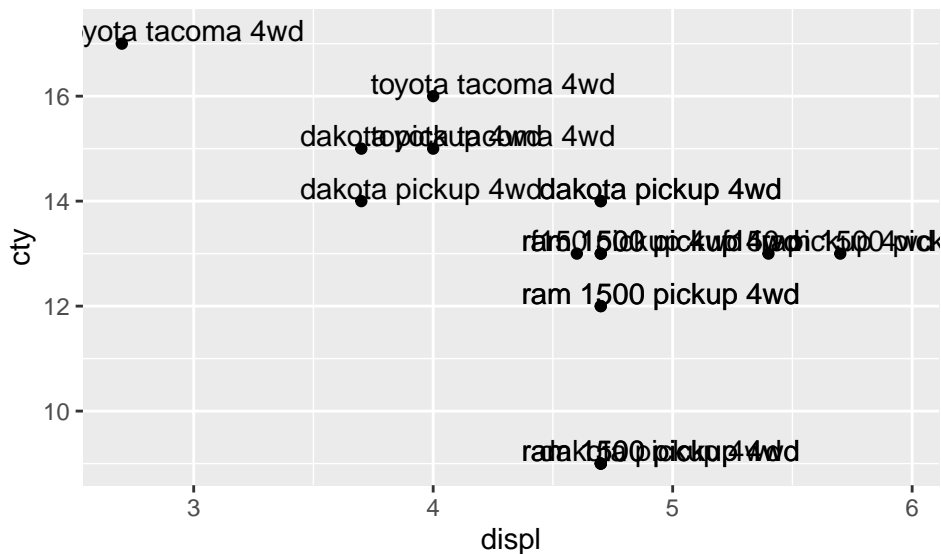


```
library(ggrepel)
mpg %>%
  filter(class == "pickup") %>%
  filter(year == 2008) %>%
  ggplot(aes(x=displ, y=cty)) +
    geom_point() +
    geom_text_repel(aes(label=model))
```

```
library(ggrepel)
mpg %>%
  filter(class=="pickup") %>%
  filter(year == 2008) %>%
  mutate(labels_toyota = ifelse(manufacturer=="toyota", model, NA)) %>%
  ggplot(aes(x=displ, y=cty)) +
    geom_point() +
    geom_text_repel(aes(label=labels_toyota))
```