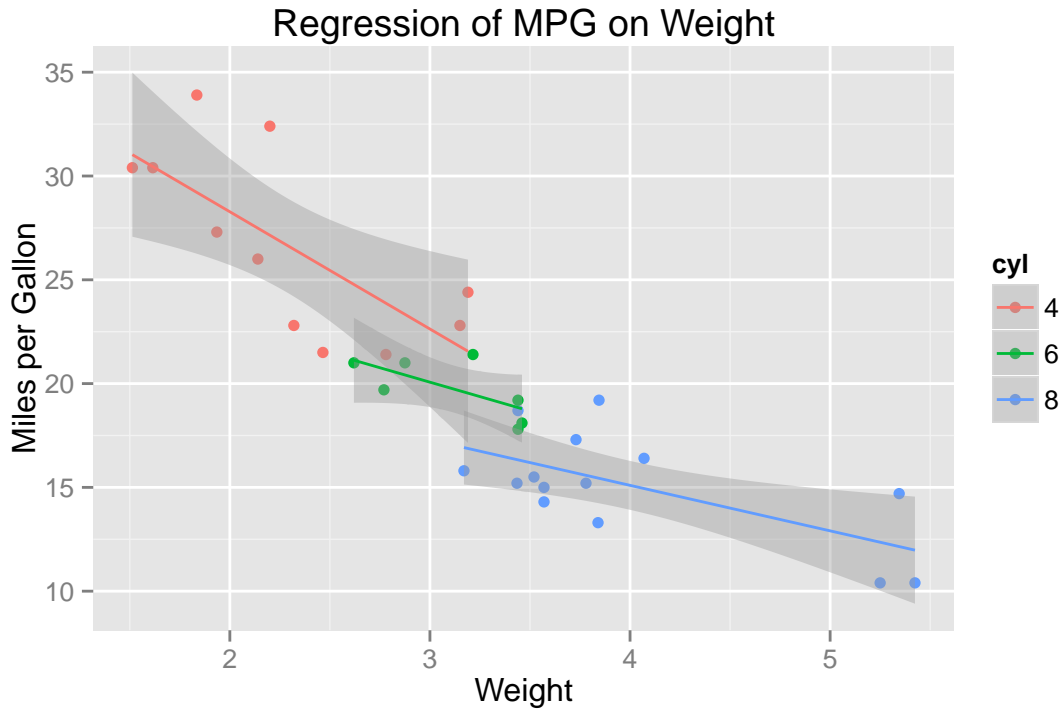# Intro to R for Epidemiologists

Lab 9 (3/19/15)

Part 1. MPG vs. Weight in mtcars dataset

The `mtcars` dataset in the `datasets` package contains fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models) from the 1974 Motor Trend magazine. Look at the help file for the `mtcars` dataset (`?mtcars`). The variables that will be used are as follows: `mpg` (miles/gallon), `cyl` (number of cylinders), and `wt` (weight lb/1000).

Using either `qplot()` or `ggplot()`, create a scatterplot of miles per gallon on weight colored by the number of cylinders. Add separate fitted lines to the plot corresponding to the number of cylinders. Note: you will need to convert the variable `cyl` to a factor.

```r
#Load data
data(mtcars)

# Separate regressions of mpg on weight for each number of cylinders
# Plot
# Set up data, x/y/ color
mtcars <- mutate(mtcars, cyl = factor(cyl))
ggplot(data = mtcars, aes(x = wt, y = mpg, colour = cyl)) +
    # Add scatterplot points
    geom_point() +
    # Add linear regression line
    geom_smooth(method = "lm") +
    # Add main and axes titles
    xlab("Weight") + ylab("Miles per Gallon") +
    ggtitle("Regression of MPG on Weight")
```

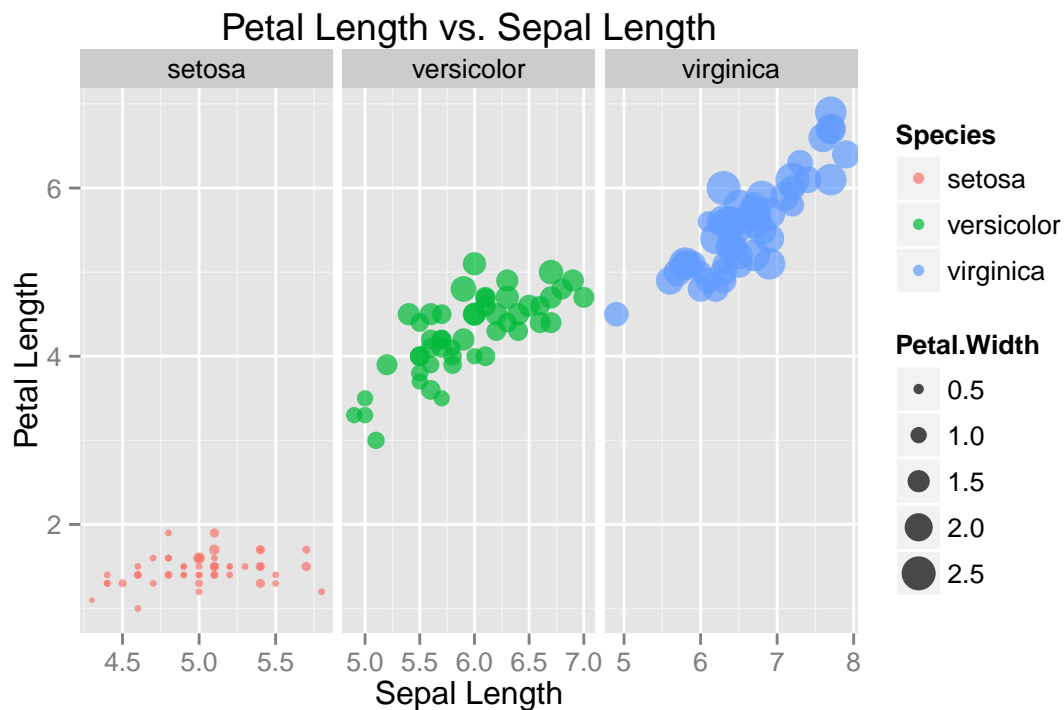Regression of MPG on Weight

```
# Alternatively using qplot
# qplot(wt, mpg, data = mtcars, geom = c("point", "smooth"),
#     method = "lm", color = cyl,
#     main = "Regression of MPG on Weight",
#     xlab = "Weight", ylab = "Miles per Gallon")
```

Part 2. Scatterplot of Sepal.Length and Petal.Length

With the `iris` dataset in R, create a scatterplot of petal length vs. sepal length separately for each species using either `qplot()` or `ggplot()`. Use different colors for each species and let the size of each point denote petal width. Set `alpha=0.7` to reduce the effects of overplotting. Note: you may need to change the argument `scales` in the `facet_wrap` function to allow the x-axis to vary between plots (see `?facet_wrap`).

```r
data(iris)

# Plot
# Set up data, x/y/ color
ggplot(iris, aes(Sepal.Length, Petal.Length, color = Species,
    size = Petal.Width)) +
    # Add points for scatterplot
    geom_point(alpha = I(0.7)) +
    # Add main and axes titles
    ggtitle("Petal Length vs. Sepal Length") +
    xlab("Sepal Length") + ylab("Petal Length") +
    # Use facet_wrap to create separate plots for species
    facet_wrap(~ Species, scales = "free_x")
```



```r
# Note: by setting the alpha of each point to 0.7, we reduce the effects of overplotting.

# Alternatively using qplot
# qplot(Sepal.Length, Petal.Length, data = iris, color = Species,
#       size = Petal.Width, alpha = I(0.7), xlab = "Sepal Length",
#       ylab = "Petal Length", main = "Petal Length vs. Sepal Length") +
# facet_grid( ~Species, scales = "free_x")
```
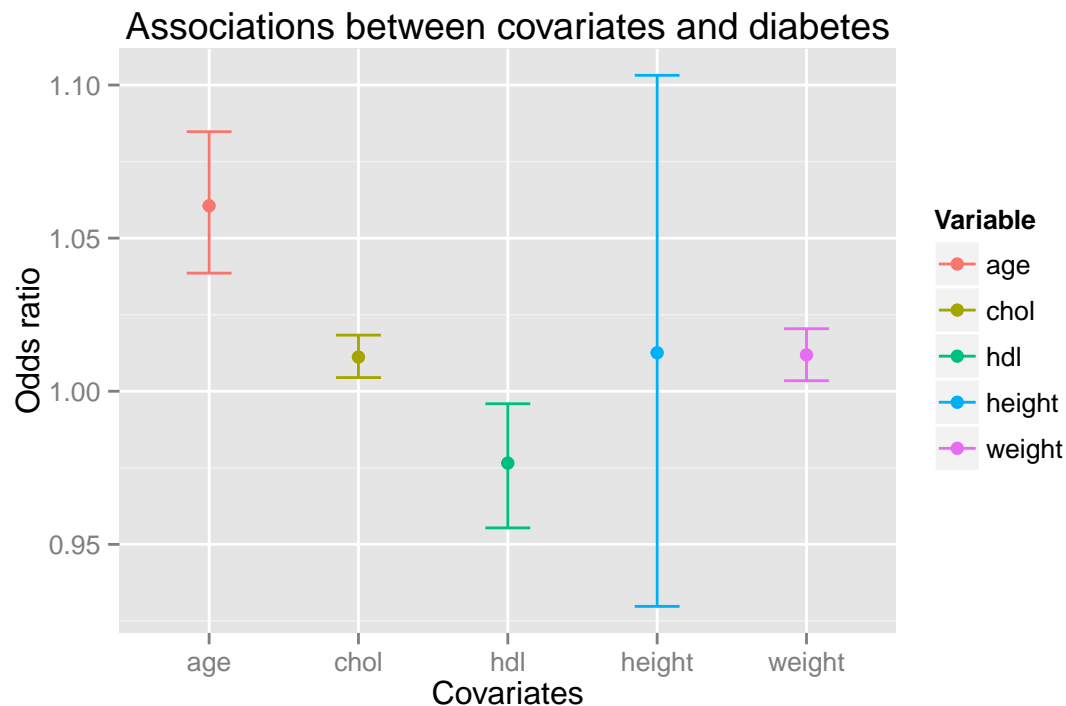
Part 3. Confidence Intervals

The dataset `OR\_df.RData` contains the odds ratios and corresponding confidence intervals from Lab 8. Use the `OR\_df` data frame to create the plot below displaying the odds ratios and corresponding confidence intervals.

```r
# Read in data
load("OR_df.RData")

#plot
# Set up data, x/y/ color
ggplot(OR_df, aes(x = Variable, y = OR, color = Variable)) +
  #Add points for ORs
  geom_point(size = 3, shape = 20) +
  #Add error bars
  geom_errorbar(aes(ymin = LB, ymax = UB), width = 0.3) +
  #Add main and axes titles
  ggtitle("Associations between covariates and diabetes") +
  ylab("Odds ratio") + xlab("Covariates")
```
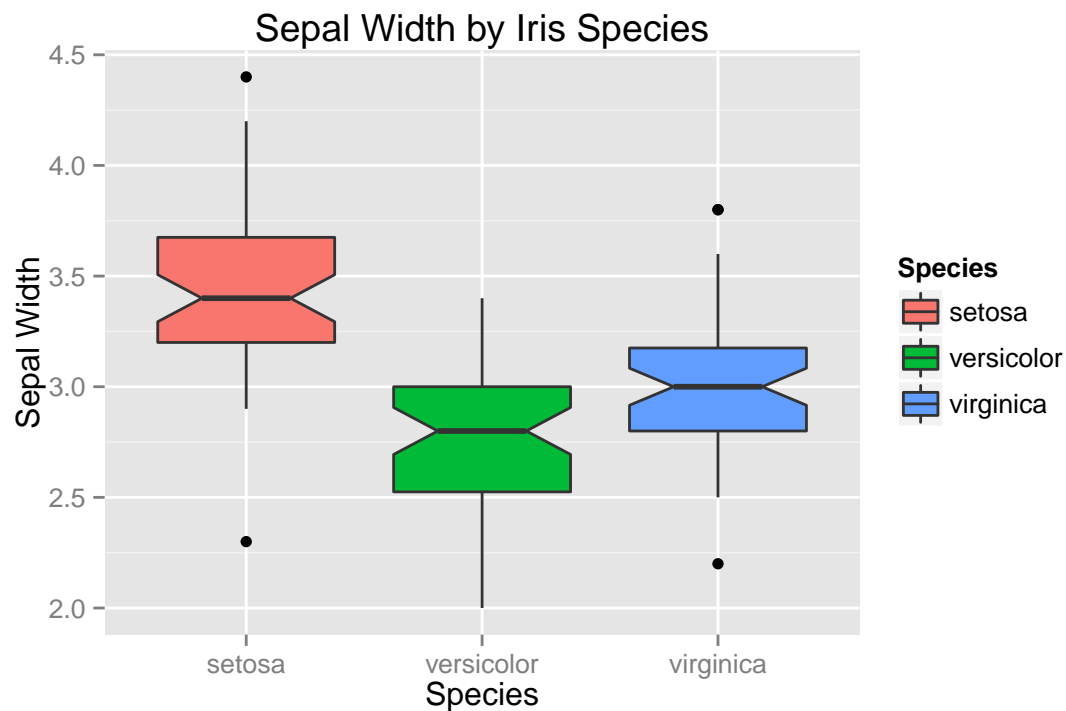


```r
# Alternatively using qplot
# qplot(Variable, OR, data = OR_df, color = Variable, geom = c("point",
#   "errorbar"), ymin = LB, ymax = UB, width = 0.3,
#   ylab = "Odds ratio", xlab = "Covariates",
#   main = "Associations between covariates and diabetes")
```

Part 4. Boxplot of sepal width by species

Use the `iris` dataset in R to create notched boxplots of sepal width by species. Make each box a different color. To color the interior of the boxplots, specify the `fill` argument instead of `colour`. Note: to create notched boxplots, you may want to consult the help page for `geom_boxplot` (`?geom_boxplot`).

```r
# Plot
# Set up data, x/y/ color (fill)
ggplot(iris, aes(Species, Sepal.Width, fill = Species)) +
  #Add notched boxplot
  geom_boxplot(notch = TRUE) +
  # Add main and axes titles
  ggtitle("Sepal Width by Iris Species") +
  xlab("Species") + ylab("Sepal Width")
```



```r
# # Alternatively using qplot
# qplot(Species, Sepal.Width, data = iris, geom = "boxplot", notch = T,
#   fill = Species, xlab = "Species", ylab = "Sepal Width",
#   main = "Sepal Width by Iris Species")
```