```r
# Video 2 - Reading in the Dataset

# Get the current directory
  getwd()
# Read the csv file
  USDA = read.csv("USDA.csv")
# Structure of the dataset
  str(USDA)
# Statistical summary
  summary(USDA)


# Video 3 - Basic Data Analysis

# Vector notation
  USDA$Sodium
# Finding the index of the food with highest sodium levels
  which.max(USDA$Sodium)
# Get names of variables in the dataset
  names(USDA)
# Get the name of the food with highest sodium levels
  USDA$Description[265]
# Create a subset of the foods with sodium content above 10,000mg
  HighSodium = subset(USDA, Sodium>10000)
# Output names of the foods with high sodium content
  HighSodium$Description
# Finding the index of CAVIAR in the dataset
  match("CAVIAR", USDA$Description)
# Find amount of sodium in caviar
  USDA$Sodium[4154]
# Doing it in one command!
  USDA$Sodium[match("CAVIAR", USDA$Description)]
# Summary function over Sodium vector
  summary(USDA$Sodium)
# Standard deviation
  sd(USDA$Sodium, na.rm = TRUE)


# Video 4 - Plots

# Scatter Plots
  plot(USDA$Protein, USDA$TotalFat)
# Add xlabel, ylabel and title
  plot(USDA$Protein, USDA$TotalFat, xlab="Protein", ylab = "Fat",
 main = "Protein vs Fat", col = "red")
# Creating a histogram
  hist(USDA$VitaminC, xlab = "Vitamin C (mg)", main = "Histogram of
 Vitamin C")
```

```r
# Add limits to x-axis
  hist(USDA$VitaminC, xlab = "Vitamin C (mg)", main = "Histogram of
 Vitamin C", xlim = c(0,100))
# Specify breaks of histogram
  hist(USDA$VitaminC, xlab = "Vitamin C (mg)", main = "Histogram of
 Vitamin C", xlim = c(0,100), breaks=100)
  hist(USDA$VitaminC, xlab = "Vitamin C (mg)", main = "Histogram of
 Vitamin C", xlim = c(0,100), breaks=2000)
# Boxplots
  boxplot(USDA$Sugar, ylab = "Sugar (g)", main = "Boxplot of Sugar")


# Video 5 - Adding a variable

# Creating a variable that takes value 1 if the food has higher
 sodium than average, 0 otherwise
  HighSodium = as.numeric(USDA$Sodium > mean(USDA$Sodium,
 na.rm=TRUE))
  str(HighSodium)
# Adding the variable to the dataset
  USDA$HighSodium = as.numeric(USDA$Sodium > mean(USDA$Sodium,
 na.rm=TRUE))
# Similarly for HighProtein, HigCarbs, HighFat
  USDA$HighCarbs = as.numeric(USDA$Carbohydrate > mean(USDA
 $Carbohydrate, na.rm=TRUE))
  USDA$HighProtein = as.numeric(USDA$Protein > mean(USDA$Protein,
 na.rm=TRUE))
  USDA$HighFat = as.numeric(USDA$TotalFat > mean(USDA$TotalFat,
 na.rm=TRUE))


# Video 6 - Summary Tables

# How many foods have higher sodium level than average?
  table(USDA$HighSodium)
# How many foods have both high sodium and high fat?
  table(USDA$HighSodium, USDA$HighFat)
# Average amount of iron sorted by high and low protein?
  tapply(USDA$Iron, USDA$HighProtein, mean, na.rm=TRUE)
# Maximum level of Vitamin C in hfoods with high and low carbs?
  tapply(USDA$VitaminC, USDA$HighCarbs, max, na.rm=TRUE)
# Using summary function with tapply
  tapply(USDA$VitaminC, USDA$HighCarbs, summary, na.rm=TRUE)
```