

Navajo Nation Water Quality

Aishwarya

January 31, 2019

```
#install.packages("maps")
#install.packages("mapproj")
#install.packages("measurements")
```

Mutate the dataset to replace the negative values of Radium-228 with 0, then filter the dataset to remove any sites with “Unknown Risk” for the EPA risk rating. Visualize the distribution of Radium-228 within each EPA section and each risk level

```
data <- read_csv("C:/Users/Aishwarya/Desktop/NEU/Introduction to Data management/hw2/NavajoWater.csv" )
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `Amount of Aluminum (Al)` = col_number(),
##   `Amount of Antimony (Sb)` = col_double(),
##   `Amount of Arsenic (As)` = col_double(),
##   `Amount of Barium (Ba)` = col_number(),
##   `Amount of Beryllium (Be)` = col_double(),
##   `Amount of Cadmium (Cd)` = col_double(),
##   `Amount of Chromium (Cr)` = col_double(),
##   `Amount of Copper (Cu)` = col_double(),
##   `Amount of Iron (Fe)` = col_number(),
##   `Amount of Lead (Pb)` = col_double(),
##   `Amount of Manganese (Mn)` = col_number(),
##   `Amount of Mercury (Hg)` = col_double(),
##   `Amount of Nickel (Ni)` = col_double(),
##   `Amount of Selenium (Se)` = col_double(),
##   `Amount of Silver (Ag)` = col_double(),
##   `Amount of Thallium (TI)` = col_double(),
##   `Amount of Vanadium (V)` = col_double(),
##   `Amount of Zinc (Zn)` = col_number(),
##   `Amount of Alpha Particles` = col_double(),
##   `Amount of Beta Particles` = col_double()
##   # ... with 9 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
#Removing Negative Values in Amount of Radium228 Coloumn
old_col <- data$`Amount of Radium228`
new_col <- ifelse(old_col < 0, 0, old_col)
data <- mutate(data, 'Amount of Radium228' = new_col)
```

```
#Printing head
head(data$`Amount of Radium228`)
```

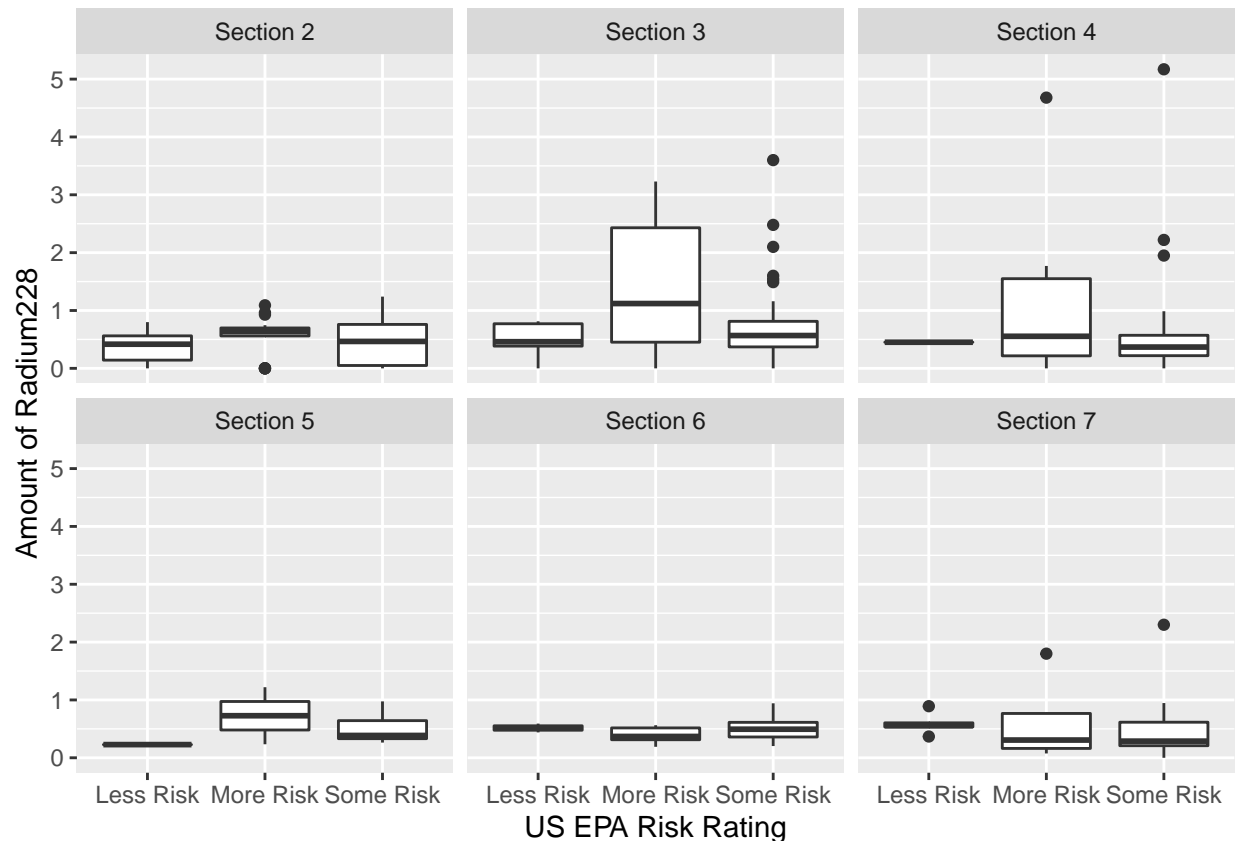
```
## [1] 0.500 1.540 0.591 0.183 0.439 0.892
```

```
#Removing row containing 'Unknown Risk'
```

```
data <- data[-(which(data$`US EPA Risk Rating` == 'Unknown Risk')), ]
```

```
#Plotting Boxplots
```

```
ggplot(data = data, mapping = aes( x = `US EPA Risk Rating`, y = `Amount of Radium228`)) + geom_boxplot()
```



OBSERVATION:-

Section 3,4,7 have comparatively more amount of radium228 which are of higher risk than other sections.

Create a map of the region (you may want to adjust the plotting limits to an appropriate “zoom” level) showing the locations of the water sampling sites, along with the EPA risk and the concentration of Radium-228 for each location

```
#PLOTING FOUR CORNERS
```

```
four_corners <- ggplot2::map_data("state", region = c("arizona", "New Mexico", "Utah", "Colorado"))
```

```
#CONVERTING LAT AND LONG DATA
```

```
data$Longitude <- measurements::conv_unit(data$Longitude, "deg_min_sec", "dec_deg")
```

```
data$Latitude <- measurements::conv_unit(data$Latitude, "deg_min_sec", "dec_deg")
```

```
#PLOTING POINTS ON FOUR CORNERS
```

```
ggplot(four_corners) +  
  geom_polygon(mapping=aes(x=long,  
                           y=lat,
```

```

        group = group),
      fill=NA,
      color="black") +
    geom_point(data, mapping = aes(x = -as.double(Longitude),
                                   y = as.double(Latitude),
                                   shape = `US EPA Risk Rating`,
                                   color = `Amount of Radium228`)) +
    #ZOOMING
    coord_fixed(xlim = c(-112.1, -109.1), ylim = c(35.1,37.1), ratio = 1.3) +
    ggtitle("Locations of Water Sampling sites")

```

