Mario\_Ruiz\_HW6

Mario Ruiz

2/25/2020

## Load Airquality data set

my.data <- airquality my.data

## Remove NAs - first see what NAs are available

colSums(is.na(my.data)) ## Remove NAs cleandata <- na.omit(my.data) ## Verify colSums(is.na(cleandata))

## Histograms for each variable

library(ggplot2) ozoneViz <- ggplot(cleandata, aes(x=Ozone)) + geom\_histogram(bins = 10, color=“black”, fill=“red”) ozoneViz + ggtitle(“Ozone Information”)

windViz <- ggplot(cleandata, aes(x=Wind)) + geom\_histogram(bins = 10, color=“black”, fill=“blue”) windViz + ggtitle(“Wind Information”)

solarViz <- ggplot(cleandata, aes(x=Solar.R)) + geom\_histogram(bins = 10, color=“black”, fill=“yellow”) solarViz + ggtitle(“Solar Information”)

tempViz <- ggplot(cleandata, aes(x=Temp)) + geom\_histogram(bins = 10, color=“black”, fill=“green”) tempViz + ggtitle(“Temp Information”)

monthViz <- ggplot(cleandata, aes(x=Month)) + geom\_histogram(bins = 12, color=“black”, fill=“orange”) monthViz + ggtitle(“Month Information”)

dayViz <- ggplot(cleandata, aes(x=Day)) + geom\_histogram(bins = 31, color=“black”, fill=“purple”) dayViz + ggtitle(“Day Information”)

## Boxplot for Ozone

ozoneBox <- ggplot(cleandata, aes(x=factor(0), y=Ozone)) + geom\_boxplot() + coord\_flip() ozoneBox + ggtitle(“Ozone Boxplot”)

##Boxplot for Wind windBox <- ggplot(cleandata, aes(x=factor(0), y=Wind)) + geom\_boxplot() + coord\_flip() windBox + ggtitle(“Wind Boxplot”)

## Dated since 1973

dated <- paste(“1973”, cleandataDay, sep = “-”) ## Sanity check dated ## Format date dated.v2 <- as.Date(dated) dated.v2 ## Revise with newly formated date dfdated.v2 <- cbind(cleandata, dated.v2) dfdated.v2

## Line chart for ozone

ozLine <- ggplot(dfdated.v2, aes(x=dated.v2, y=Ozone)) + geom\_line(size=2, color=“green”) ozLine + ggtitle(“Ozone Line Chart”)

## Line chart for temp

teLine <- ggplot(dfdated.v2, aes(x=dated.v2, y=Temp)) + geom\_line(size=2, color=“red”) teLine + ggtitle(“Temp Line Chart”)

## Line chart for wind

wiLine <- ggplot(dfdated.v2, aes(x=dated.v2, y=Wind)) + geom\_line(size=2, color=“blue”) wiLine + ggtitle(“Wind Line Chart”)

## Line chart for solar.r

ozLine <- ggplot(dfdated.v2, aes(x=dated.v2, y=Solar.R)) + geom\_line(size=2, color=“black”) ozLine + ggtitle(“Solar.R Line Chart”)

## Combine all into one chart

## Normalize - Min/Max

normOz <- (cleandataOzone))/(max(cleandataOzone)) normOz normTe <- (cleandataTemp))/(max(cleandataTemp)) normTe normWi <- (cleandataWind))/(max(cleandataWind)) normWi normSo <- (cleandataSolar.R))/(max(cleandataSolar.R)) normSo ## Create new df with normalized data dfNorm <- data.frame(normOz, normTe, normWi, normSo, “Date”=dfdated.v2$dated.v2, "Month"=dfdated.v2$Month, “Day”=dfdated.v2$Day) ## Plot it everything <- ggplot(dfNorm, aes(x=dated.v2)) + geom\_line(aes(y=normOz), size=1, color="blue") + geom\_line(aes(y=normTe), size=1, color="red") + geom\_line(aes(y=normWi), size=1, color="purple") + geom\_line(aes(y=normSo), size=1, color="green") everything + ggtitle("Normalized Everything") + ylab("Normalized via Min/Max") ## Heatmap hmDates <- c(dfdated.v2$dated.v2, dfdated.v2dated.v2, dfdated.v2Day, dfdated.v2Day, dfdated.v2normOz, dfNormnormWi, dfNorm$normSo) hmCat <- c(rep(“Ozone”,111), rep(“Temp”,111), rep(“Wind”,111), rep(“Solar.R”,111)) dfHM <- data.frame(hmDates,hmDay,hmVars,hmCat) hm <- ggplot(dfHM, aes(x-hmDay, y=hmCat, fill=hmVars)) + geom\_tile() ## ## FIX hm + scale\_fill\_gradient(name = “Normal”, low =“green”, high = “yellow”) + ggtitle(“Air Quality Heat Map”) + ylab(“Category”) + xlab(“Day”) ##

scatData <- ggplot(dfHM, aes(x = normWi, y = normTe )) + geom\_point(aes(size=normOz, color=normSo)) scatData + scale\_color\_gradient(low = “blue”, high = “orange”)

### SCATTER PLOT FOR THE WIN!