**Day 4 Assignment**

Q. Explore ggplot2 library

**Code:**

#Exploring geom attribute

qplot(gear,mpg,data=mtcars, geom=c("point"), fill=gear, main="Mileage by Gear Number", xlab="", ylab="Miles per Gallon")

qplot(gear,mpg,data=mtcars, geom=c("smooth"), fill=gear, main="Mileage by Gear Number", xlab="", ylab="Miles per Gallon")

qplot(gear,mpg,data=mtcars, geom=c("smooth", "jitter"), fill=gear, main="Mileage by Gear Number", xlab="", ylab="Miles per Gallon")

qplot(gear,mpg,data=mtcars, geom=c("line", "smooth"), fill=gear, main="Mileage by Gear Number", xlab="", ylab="Miles per Gallon")

qplot(gear,mpg,data=mtcars, geom=c("jitter"), fill=gear, main="Mileage by Gear Number", xlab="", ylab="Miles per Gallon")

# Just exploring various options in ggplot2

ggplot(mtcars, aes(x = mtcars$mpg)) + geom\_histogram()

#defining binwidth (large bins)

ggplot(mtcars, aes(x = mtcars$mpg)) + geom\_histogram(binwidth=2)

# defining small bins

ggplot(mtcars, aes(x = mtcars$mpg)) + geom\_histogram(binwidth=0.1)

ggplot(mtcars, aes(x = mtcars$mpg)) + geom\_density(fill="darkgrey")

# Violin plots

p <- ggplot(mpg, aes(x=factor(cyl), y=hwy, fill=factor(cyl)))

p + geom\_violin(scale = "width")

p + geom\_violin(scale = "width") + geom\_point(size=2, position="jitter")

#Scatter plot

gg <- ggplot(midwest, aes(x=area, y=poptotal)) +

geom\_point(aes(col=state, size=popdensity)) +

geom\_smooth(method="loess", se=F) +

xlim(c(0, 0.1)) +

ylim(c(0, 500000)) +

labs(subtitle="Area Vs Population",

y="Population",

x="Area",

title="Scatterplot",

caption = "Source: midwest")

plot(gg)

pd <- ggplot(data = diamonds, aes(carat, price))

pd + geom\_point()

pd+ geom\_point(alpha=0.2)

qplot(mpg, data=mtcars, geom="density", fill=gear, alpha=I(.5),

main="Distribution of Gas Milage", xlab="Miles Per Gallon",

ylab="Density")

# Separate regressions of mpg on weight for each number of cylinders

qplot(wt, mpg, data=mtcars, geom=c("point", "smooth"),

method="lm", formula=y~x, color=cyl,

main="Regression of MPG on Weight",

xlab="Weight", ylab="Miles per Gallon")

p <- ggplot(mtcars, aes(x = hp, y = mpg)) + geom\_point()

p + stat\_smooth(method = "lm", formula = y ~ x, size = 1)

p + stat\_smooth(method = "lm", formula = y ~ x + I(x^2), size = 1)

require(mgcv)

p + stat\_smooth(method = "gam", formula = y ~ s(x), size = 1)

p + stat\_smooth(method = "gam", formula = y ~ s(x, k = 3), size = 1)

## if we wanted the points coloured, but not separate lines there are two

## options---force stat\_smooth() to have one group

ggplot(mtcars, aes(x = hp, y = mpg, colour = factor(vs))) + geom\_point() +

stat\_smooth(aes(group = 1), method = "lm", formula = y ~ x, se = FALSE)

## or only add colour to the points, not in the global ggplot() call

ggplot(mtcars, aes(x = hp, y = mpg)) + geom\_point(aes(colour = factor(vs))) +

stat\_smooth(method = "lm", formula = y ~ x, se = FALSE)