library(dplyr)

library(gridExtra)

library(tools)

library(zipcode)

library(maptools)

library(rgeos)

library(ggplot2)

library(scales)

library(reshape2)

library(ggmap)

library(ggthemes)

library(RColorBrewer)

library(broom)

library(rgdal)

library(corrplot)

library(stringr)

dark.blue <- "#3366cc"

Call.Data <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/311\_Call\_Center\_Tracking\_Data.csv")

#Call.Data <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/311\_Call\_Center\_Tracking\_Data.csv")

#C:\Users\453831\Documents\dso545\Project

c.d <- Call.Data %>% filter(!is.na(Zip.Code), Zip.Code <= 99950 , Zip.Code >= 90000) #99950 is highest zip code

#nrow(c.d)/nrow(Call.Data)

#summary(c.d$Zip.Code)

c.d %>% arrange(-Zip.Code) %>% head()

ZIP.income <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/ZIP-income.csv")

zcta\_county\_rel\_10 <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/zcta\_county\_rel\_10.csv")

gov <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/US Zip Codes from 2013 Government Data.csv")

names(ZIP.income)[1] <- "ZIP"

names(zcta\_county\_rel\_10)[1] <- "ZIP"

m <- merge(zcta\_county\_rel\_10, gov, by="ZIP") #had more records than full\_join

m2 <- full\_join(ZIP.income, m, by = "ZIP") #had more records than merge

m3 <- m2 %>% filter(ZIP %in% unique(c.d$Zip.Code))

c.d <- c.d %>% rename(ZIP = Zip.Code)

m4 <- left\_join(c.d, m3, by="ZIP")

m5 <- m4 %>% filter(STATE == 6)

names(m5)[1] <- "Date"

#enriched data contains ~85% of original rows

1-nrow(m5)/nrow(Call.Data)

#verifying against zip code package

data(zipcode)

#Count of entries without income figures.

m5 %>% filter(is.na(Median)) %>% select(ZIP) %>% count()

#list of zips without income figures, most are non-residential

#m5 %>% filter(is.na(Median)) %>% select(ZIP)

head(c.d)

m6 <- merge(c.d, zipcode, by.x = "ZIP", by.y = "zip")

summary(m6)

#verifying sanity of geo-data

summary(m6$longitude - m6$LNG)

summary(m6$latitude - m6$LAT)

#list of cities

unique(m6$city)

colors <- brewer.pal(9, "BuGn")

county <- readShapePoly("C:/Users/Chris/Desktop/DSO-545/Project/cb\_2015\_us\_county\_500k.shp")

area <- readShapePoly("C:/Users/Chris/Desktop/DSO-545/Project/cb\_2015\_us\_zcta510\_500k.shp")

area2 <- area[area$ZCTA5CE10 %in% unique(m6$ZIP),]

area.points <- fortify(area2)#, region=area2$ZCTA5CE10)

#I needed to chop off the region part

county2 <- county

county2$STATEFP <- as.numeric(as.character(county2$STATEFP))

summary(county2$STATEFP)

county3 <- county2[(county2$STATEFP %in% 6),]

county.points <- fortify(county3)

mapImage <- get\_map(location = c(lon = -118, lat = 34), color = "color", maptype = "watercolor", zoom = 8)

states <- map\_data("state")

california <- subset(states, region %in% c("california"))

county <- map\_data("county")

la.counties <- subset(county, region %in% c("california"))

###baselineplot

#ggmap(mapImage) +

#older/darker image template

#ggplot()+

# geom\_polygon(data = california, aes(x = long, y = lat, group = group), color = "white", fill = "#009900") +

# geom\_polygon(aes(x = long, y = lat, group = group), data = area.points, color = colors[9], fill = colors[6], alpha = 1) + #coordmap limits prevent clipping

# geom\_polygon(aes(x = long, y = lat, group = group), data = county.points, color = "white", size = .8, alpha = .8, fill=NA) + #coordmap limits prevent clipping

# coord\_map(projection="mercator", xlim=c(-119.5 , -116.5), ylim=c(33.4,34.85)) +

# theme\_map() + theme(panel.background = element\_rect(fill = "#3399ff"))

p <- ggplot() +

geom\_polygon(data = california, aes(x = long, y = lat, group = group), color = "white", fill = "#d3d3d3") +

geom\_polygon(aes(x = long, y = lat, group = group), data = area.points, color = "#a0a0a0", fill = "#d3d3d3", alpha = 1) + #coordmap limits prevent clipping

geom\_polygon(aes(x = long, y = lat, group = group), data = county.points, color = "#dddddd", size = .9, alpha = .8, fill = NA) + #coordmap limits prevent clipping

xlim(-119.5 - 1, -116.5 + 2.4) + ylim(33.4 - .95, 34.85 + 2) + #this is added to reduce needless drawing off screen

coord\_map(projection = "mercator", xlim = c(-119.5 , -116.5), ylim = c(33.4,34.85)) + #this reduces what is show on screne

theme\_map() + theme(panel.background = element\_rect(fill = "#96ccfd"))

p

#sample with densities - runs slowly

#p + stat\_density2d(data = m6[sample(1:nrow(m6),round(nrow(m6)/100)),],

# aes(x = jitter(longitude), y = jitter(latitude), alpha = ..level..),

# fill = "black",

# color = NA,

# geom = "polygon")

#sample with bins - runs quickly

#png("C:/Users/Chris/Desktop/DSO-545/Project/MapTotal.png", width=16, height=8, units="in", res=300)

m6$Day <- m6$Date %>% mdy() %>% wday(label= TRUE, abbr=FALSE)

head(m6$Day)

png("C:/Users/Chris/Desktop/DSO-545/Project/MapWday.png", width=16, height=8, units="in", res=300)

p + geom\_bin2d(data = m6,#[sample(1:nrow(m6), round(nrow(m6)/1000)),],

aes(x = jitter(longitude), y = jitter(latitude)),

binwidth = c(0.05, 0.05)) +

scale\_fill\_gradient(low = "#3399FF", high = "#003399") +

ggtitle("Total Requests") + theme(legend.justification = c(1, 0), legend.position = c(1, 0)) +

facet\_wrap( ~ Day)

dev.off()

m6x <- m6 %>% group\_by(Service.Name) %>% summarise(serv.count = n()) %>% ungroup() %>% arrange(-serv.count)

m6x$S.Name <- "Other"

m6x$S.Name[1:5] <- as.character(unlist(m6x$Service.Name))

m6x <- merge(m6x, m6, by="Service.Name")

m6x$S.Name <- as.factor(m6x$S.Name)

summary((m6x$S.Name)) %>% arrange(summary((m6x$S.Name)))

summary((m6x$S.Name))

m6x$S.Name <- factor(m6x$S.Name, levels = c("Ideal", "Very Good", "Fair", "Good", "Premium"))

#unique(m6x$S.Name) %in% c("Bulky Item Pick-up","Online Request for Permit Inspection","","Graffiti Removal - Community Beautification", "Subject Specialty Group","Other")

m6x$S.Name <- factor(m6x$S.Name, levels = c("Bulky Item Pick-up","Online Request for Permit Inspection","","Graffiti Removal - Community Beautification", "Subject Specialty Group","Other"))

png("C:/Users/Chris/Desktop/DSO-545/Project/MapService.png", width=16, height=8, units="in", res=300)

p + geom\_bin2d(data = m6x,#[sample(1:nrow(m6), round(nrow(m6)/1000)),],

aes(x = jitter(longitude), y = jitter(latitude)),

binwidth = c(0.05, 0.05)) +

scale\_fill\_gradient(low = "#3399FF", high = "#003399") +

ggtitle("Total Requests") + theme(legend.justification = c(1, 0), legend.position = c(1, 0)) +

facet\_wrap( ~ S.Name)

dev.off()

Department.Name <- unique(m6$Department.Name)

###

### Plot of top 20 Requests

###

d1 <- m6 %>% group\_by(Service.Name) %>% summarise(count=n()) %>%

mutate(Service.Name = toTitleCase(as.character(Service.Name))) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, " - ", "\n")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, " \\(", "\n\\(")) %>% mutate(Count = count/1000) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, " / ", "/")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, " ", "\n")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, "a\n", "a ")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, "of\n", "of ")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, "for\n", "for ")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, "and\n", "& ")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, "\\+\n", "+ ")) %>%

mutate(Service.Name = str\_replace\_all(Service.Name, "\n&", " &")) %>%

arrange(-count) %>% slice(1:20) %>% mutate(Var = c(rep(1, 10), rep(2, 10)))

d1

Calls1 <- d1 %>% slice(1:10) %>% ggplot(aes(x = reorder(Service.Name, -count), y = Count)) + geom\_bar(stat = "identity", fill = dark.blue) +

ylab("Count (thousands)") + xlab("") + theme\_minimal() + ylim(0, 600)

Calls2 <- d1 %>% slice(11:20) %>% ggplot(aes(x = reorder(Service.Name, -count), y = Count)) + geom\_bar(stat = "identity", fill = dark.blue) +

ylab("Count (thousands)") + xlab("") + theme\_minimal() + ylim(0, 600)

png("C:/Users/Chris/Desktop/DSO-545/Project/requests.png", width=16, height=8, units="in", res=300)

grid.arrange(Calls1, Calls2, top = "Top Requests")

dev.off()

###

### Plot of Top 20 Departments

###

d2 <- m6 %>% group\_by(Department.Name) %>% summarise(count=n()) %>%

mutate(Department.Name = toTitleCase(as.character(Department.Name))) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, " - ", "\n")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, " \\(", "\n\\(")) %>% mutate(Count = count/1000) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, " / ", "/")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, " ", "\n")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, "a\n", "a ")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, "of\n", "of ")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, "for\n", "for ")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, "and\n", "& ")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, "\n\\+", " +")) %>%

mutate(Department.Name = str\_replace\_all(Department.Name, "\n&", " &")) %>%

arrange(-count) %>% slice(1:20) %>% mutate(Var = c(rep(1, 10), rep(2, 10)))

Calls3 <- d2 %>% slice(1:10) %>% ggplot(aes(x = reorder(Department.Name, -count), y = Count)) + geom\_bar(stat = "identity", fill = dark.blue) +

ylab("Count (thousands)") + xlab("") + theme\_minimal() + ylim(0, 1100)

Calls4 <- d2 %>% slice(11:20) %>% ggplot(aes(x = reorder(Department.Name, -count), y = Count)) + geom\_bar(stat = "identity", fill = dark.blue) +

ylab("Count (thousands)") + xlab("") + theme\_minimal() + ylim(0, 1100)

png("C:/Users/Chris/Desktop/DSO-545/Project/depts.png", width=16, height=8, units="in", res=300)

grid.arrange(Calls3, Calls4, top = "Calls to each Department")

dev.off()

m6$Department.Name <- as.character(m6$Department.Name)

m6$Service.Name <- as.character(m6$Service.Name)

head(m6)

census <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/zcta\_master.csv")[,c(2,70,73:95)]

#census <- read.csv("C:/Users/Chris/Desktop/DSO-545/Project/zcta\_master.csv")[,c(2,70,73:95,25,26,66)]

census$ZIP <- census$zcta5

#names(census) == "TotPopACS"

###DEFINE A PLOT FUNCTION based on indices

#for plotting frequency normalized to population

myplot <- function(dx){

ggplot(dx, aes\_string(colnames(dx)[1], colnames(dx)[2])) + geom\_bin2d() + geom\_smooth(se=FALSE) + theme\_minimal() + scale\_fill\_gradient(low = "grey", high = "black")

}

#for plotting histograms on raw counts

myplot2 <- function(dx){

# ggplot(dx[,1]) + geom\_histogram(weight=dx[,2]) + theme\_minimal() + scale\_fill\_gradient(low = "grey", high = "black")

qplot(dx[,1], weight=dx[,2], geom="histogram") + theme\_minimal() + geom\_histogram(fill = dark.blue, color="#2255dd")

}

myplot3 <- function(dx){

ggplot(dx, aes\_string(colnames(dx)[1], colnames(dx)[2])) + geom\_point(color="blue", alpha=.2, size=3, fill=dark.blue, pch=21) + geom\_smooth(se=FALSE, color="black") + theme\_minimal()

}

dark.blue

i=1

###begin loop here

fd <- m6 %>% group\_by(ZIP) %>% summarise(count = n())

full.data <- left\_join(fd, census, by = "ZIP") %>% arrange(-count)

write.csv(full.data, "TopZips.csv")

#Plot of Top Zip Codes

png("C:/Users/Chris/Desktop/DSO-545/Project/TopZips.png", width=16, height=8, units="in", res=300, type = "cairo")

full.data[,1:2] %>% slice(1:10) %>% ggplot(aes(x = reorder(factor(ZIP), -count), y = count)) + geom\_bar(stat="identity", fill=dark.blue) + theme\_minimal() +

xlab("Zipcode") + ylab("Call Count") + ggtitle("Top Zipcodes by Call Frequency") #+ scale\_x\_continuous(limits = c(0, 13000), breaks=0:13\*1000)

dev.off()

for(i in 1:2) {

i <- 1

i <- 1 + i

m6i <- m6 %>% group\_by(Service.Name) %>% summarise(c = n()) %>% arrange(-c) %>% slice(i) %>% select(Service.Name) %>% unlist() %>% as.character()

m7 <- m6 %>% filter(Service.Name == m6i) %>% group\_by(ZIP) %>% summarise(count = n())

m8 <- left\_join(m7, census, by = "ZIP")

#head(m8[,sapply(m8, is.numeric)])

#cor(m8[,sapply(m8, is.numeric)], use="pairwise.complete.obs")

#plot((cor(m8[,sapply(m8, is.numeric)], use="pairwise.complete.obs")))

#plot(cor(m8[,sapply(m8, is.numeric)]))

#nums <- sapply(m8, is.numeric)

#ggplot(data = m8, aes(x = count/TotPopACS, y = TotPopACS)) + geom\_bin2d() + geom\_smooth() + theme\_minimal() + scale\_fill\_gradient(low = "grey", high = "black") + xlim(-0.1,4)

plot\_list = list()

for(j in 1:27) {

dx <- data.frame(cbind(m8[,j], Dens = m8$count/m8$TotPopACS)); dx <- dx[complete.cases(dx),]

plot\_list[[j]] = myplot(dx)

}

for(j in 1:27) {

png(str\_replace\_all(paste0("C:/Users/Chris/Desktop/DSO-545/Project/Set2/",names(m8)[j],m6i,".png"),"[[:space:]]",""), width = 16, height = 8, units = "in", res = 300)

print(plot\_list[[j]])

dev.off()

}

}

head(full.data)

#percent over 65 - Overall

over65rel <- data.frame(cbind(full.data[,5], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

over65abs <- data.frame(cbind(full.data[,5], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PercentOver65.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(over65rel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-3, 103)) + scale\_y\_continuous(limits = c(-0.05, .5), breaks=0:5/10),

myplot2(over65abs) + xlab("Percentage over 65 in ZIP Code (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1.5), breaks=0:1) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Older Zipcodes have a Higher Propensity to Call",

bottom = "Calls are somewhat Clustered around Zipcodes with around 10% of Citizens over 65")

dev.off()

#percent white - Overall

perwhiterel <- data.frame(cbind(full.data[,6], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

perwhiteabs <- data.frame(cbind(full.data[,6], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PercentWhite.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(perwhiterel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-3, 103)) + scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(perwhiteabs) + xlab("Percentage Identifying as White in ZIP Code (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with a High Proportion of Whites have a Lower Propensity to Call",

bottom = "Calls are somewhat Clustered around Zipcodes with around 60% Whites")

dev.off()

#HH Inc - Overall

HHIncrel <- data.frame(cbind(full.data[,11]/1000, Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

HHIncabs <- data.frame(cbind(full.data[,11]/1000, Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

head(HHIncabs)

png("C:/Users/Chris/Desktop/DSO-545/Project/HHIncome.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(HHIncrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-5, 230))+ scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(HHIncabs) + xlab("Household Income in ZIP Code (Thousands)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-5, 230)),

top = "Residents in Low Income Zipcodes have a Higher Propensity to Call \n Those in High Income Zipcodes have a Slightly Higher Propensity to Call",

bottom = "The Majority of Calls come from Middle Class Areas")

dev.off()

#HVal

HValrel <- data.frame(cbind(full.data[,23]/1000, Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

HValabs <- data.frame(cbind(full.data[,23]/1000, Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/HouseValue.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(HValrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-5, 1050))+ scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(HValabs) + xlab("Property Value in ZIP Code (Thousands)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-5, 1050)),

top = "Residents in Low Income Zipcodes have a Higher Propensity to Call \n Those in High Income Zipcodes have a Slightly Higher Propensity to Call",

bottom = "The Majority of Calls come from Middle Class Areas")

dev.off()

#Per Poor

PerPoorrel <- data.frame(cbind(full.data[,15], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PerPoorabs <- data.frame(cbind(full.data[,15], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerPoor.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(PerPoorrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-3, 103))+ scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(PerPoorabs) + xlab("Percentage Poor (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with many Poor have a Higher Propensity to Call",

bottom = "The Majority of Calls come from Middle Class Areas")

dev.off()

#Per in College

Percolrel <- data.frame(cbind(full.data[,17], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

Percolabs <- data.frame(cbind(full.data[,17], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerInCollege.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(Percolrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-3, 103)) + scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(Percolabs) + xlab("Percentage in College (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with many College Students have a Higher Propensity to Call")

dev.off()

#Per College degs

Percolmorerel <- data.frame(cbind(full.data[,18], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

Percolmoreabs <- data.frame(cbind(full.data[,18], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerCollegePlus.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(Percolmorerel) +ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-3, 103)) + scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(Percolmoreabs) + xlab("Percentage with College Degree or Higher (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with many College Graduates have a Higher Propensity to Call")

dev.off()

###THIS IS BIG

#Per For Born

head(full.data)

PerForrel <- data.frame(cbind(full.data[,19], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PerForabs <- data.frame(cbind(full.data[,19], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerForeignBorn.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(PerForrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-4, 104)) + scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(PerForabs) + xlab("Percentage in Zipcode Foreign Born (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-4, 104)),

top = "Residents in Zipcodes with more Foreign Born Residents have a Higher Propensity to Call")

dev.off()

#Perc Renters

PerRentersrel <- data.frame(cbind(full.data[,22], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PerRentersabs <- data.frame(cbind(full.data[,22], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerRenters.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(PerRentersrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-4, 104)) + scale\_y\_continuous(limits = c(-0.05, 2), breaks=0:2),

myplot2(PerRentersabs) + xlab("Percentage in Zipcode Renting (%)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-4, 104)),

top = "Residents in Zipcodes with more Renters have a Higher Propensity to Call")

dev.off()

head(full.data)

#Population Density

PopDensrel <- data.frame(cbind(full.data$TotPopACS/full.data[,30]/1000, Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PopDensabs <- data.frame(cbind(full.data$TotPopACS/full.data[,30]/1000, Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PopDensity.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(myplot(PopDensrel) + ylab("Calls per Resident") + xlab("") + guides(fill=FALSE) + scale\_x\_continuous(limits = c(-2, 55)) + scale\_y\_continuous(limits = c(-0.05, 3), breaks=0:3),

myplot2(PopDensabs) + xlab("Population Density (1000x per Square Mile)") + ylab("Total Calls (Millions)") + scale\_y\_continuous(limits = c(0, 1), breaks=0:1) + scale\_x\_continuous(limits = c(-2, 55)),

top = "Residents in Zipcodes with Higher Population Densities have a Higher Propensity to Call")

dev.off()

set.seed(5689)

movies <- movies[sample(nrow(movies), 1000), ]

# Simple examples

qplot(rating, data=movies, geom="histogram")

p65

Calls3 <- d2 %>% slice(1:10) %>% ggplot(aes(x = reorder(Department.Name, -count), y = Count)) + geom\_bar(stat = "identity", fill = dark.blue) +

ylab("Count (thousands)") + xlab("") + theme\_minimal() + ylim(0, 1100)

Calls4 <- d2 %>% slice(11:20) %>% ggplot(aes(x = reorder(Department.Name, -count), y = Count)) + geom\_bar(stat = "identity", fill = dark.blue) +

ylab("Count (thousands)") + xlab("") + theme\_minimal() + ylim(0, 1100)

png("C:/Users/Chris/Desktop/DSO-545/Project/depts.png", width=16, height=8, units="in", res=300)

grid.arrange(Calls3, Calls4, top = "Calls to each Department")

dev.off()

m6 %>% group\_by()

m7 <- m6 %>% filter(Service.Name == m6i) %>% group\_by(ZIP) %>% summarise(count = n())

m8 <- left\_join(m7, census, by = "ZIP")

#Generate counts by deperatement and zip code, then arrange by top Depts/ZIPs

Dept.Summary.0 <- m6 %>% group\_by(ZIP, Department.Name) %>% summarise(Count = n())

Dept.Summary.1 <- m6 %>% group\_by(Department.Name) %>% summarise(Department.Count = n()) %>% arrange(-Department.Count)

Dept.Summary.2 <- merge(Dept.Summary.0, Dept.Summary.1, by="Department.Name")

Dept.Summary <- merge(Dept.Summary.2, census, by="ZIP") %>% arrange(-Count) %>% arrange(-Department.Count)

write.csv(Dept.Summary, "C:/Users/Chris/Desktop/DSO-545/Project/Deptartment.csv")

#Generate counts by Service and zip code, then arrange by top Depts/ZIPs

Serv.Summary.0 <- m6 %>% group\_by(ZIP, Service.Name) %>% summarise(Count = n())

Serv.Summary.1 <- m6 %>% group\_by(Service.Name) %>% summarise(Service.Count = n()) %>% arrange(-Service.Count)

Serv.Summary.2 <- merge(Serv.Summary.0, Serv.Summary.1, by="Service.Name")

Serv.Summary <- merge(Serv.Summary.2, census, by="ZIP") %>% arrange(-Count) %>% arrange(-Service.Count)

write.csv(Serv.Summary, "C:/Users/Chris/Desktop/DSO-545/Project/Service.csv")

unique(Serv.Summary$Service.Name)

dim(Serv.Summary)

#top 5 by groups...

Serv.Chart.0 <- Serv.Summary %>% mutate(rank = dense\_rank(desc(Service.Count))) %>% filter(rank <= 5) %>%

group\_by(Service.Name) %>% mutate(rank2 = dense\_rank(desc(Count))) %>% filter(rank2 <= 5)

Serv.Chart.0 %>%

ggplot(aes(x = factor(ZIP), Count), y = Count) +

geom\_bar(stat="identity", fill=dark.blue) + coord\_flip() + facet\_grid(Service.Name ~., scales = "free", space = "free") + guides(fill=FALSE)

ggplot(aes(x = factor(ZIP), Count), y = Count) + geom\_bar(stat="identity", fill=dark.blue) +

facet\_grid(.~Service.Name, scales = "free", space = "free") + guides(fill=FALSE)

Serv.Chart <- Serv.Chart.0 %>% ungroup() %>% mutate(rank3 = rank(Count))

rownames(Serv.Chart)

data.frame() %>% select(c(1:5,30:31))

#Serv.Chart.0 %>% ggplot(aes(y = reorder(reorder(factor(ZIP),-Count), rank2), x = Count)) +

Serv.Chart %>% ggplot(aes(y = reorder(as.character(ZIP), rank3), x = Count)) +

geom\_point(color=dark.blue, size=8) + facet\_grid(Service.Name ~ ., scales = "free", space = "free") +

theme(strip.text.y = element\_text(angle = 0))

class(Serv.Chart.0$Count)

(Serv.Chart.1)

%>% dim() group\_by(ZIP) %>% mutate(rank2 = dense\_rank(desc(Count))) %>% filter(rank2 <= 5) %>%

dim()

x

z = 1 + z

x <- data.frame(cbind(full.data[,z], Dens = full.data$count/1000000)) %>% filter(complete.cases(.))

names(full.data)[z]

myplot2(x)

#percent over 65 - Overall

over65abs <- data.frame(cbind(full.data[,5], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PercentOver65.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(over65abs) + xlab("Percentage over 65 in ZIP Code (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 800), breaks=0:5\*200) + scale\_x\_continuous(limits = c(-3, 50)),

top = "Residents in Older Zipcodes have a Higher Propensity to Call",

bottom = "Calls are somewhat Clustered around Zipcodes with around 10% of Citizens over 65")

dev.off()

#percent white - Overall

perwhiterel <- data.frame(cbind(full.data[,6], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

perwhiteabs <- data.frame(cbind(full.data[,6], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PercentWhite.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(perwhiteabs) + xlab("Percentage Identifying as White in ZIP Code (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 400), breaks=0:5\*100) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with a High Proportion of Whites have a Lower Propensity to Call",

bottom = "Calls are somewhat Clustered around Zipcodes with around 60% Whites")

dev.off()

#HH Inc - Overall

HHIncrel <- data.frame(cbind(full.data[,11]/1000, Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

HHIncabs <- data.frame(cbind(full.data[,11]/1000, Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/HHIncome.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(HHIncabs) + xlab("Household Income in ZIP Code (Thousands)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 500), breaks=0:5\*100) + scale\_x\_continuous(limits = c(-5, 230)),

top = "Residents in Low Income Zipcodes have a Higher Propensity to Call \n Those in High Income Zipcodes have a Slightly Higher Propensity to Call",

bottom = "The Majority of Calls come from Middle Class Areas")

dev.off()

#HVal

HValrel <- data.frame(cbind(full.data[,23]/1000, Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

HValabs <- data.frame(cbind(full.data[,23]/1000, Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/HouseValue.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(HValabs) + xlab("Property Value in ZIP Code (Thousands)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 400), breaks=0:4\*100) + scale\_x\_continuous(limits = c(-5, 1050)),

top = "Residents in Low Income Zipcodes have a Higher Propensity to Call \n Those in High Income Zipcodes have a Slightly Higher Propensity to Call",

bottom = "The Majority of Calls come from Middle Class Areas")

dev.off()

#Per Poor

PerPoorrel <- data.frame(cbind(full.data[,15], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PerPoorabs <- data.frame(cbind(full.data[,15], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerPoor.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(PerPoorabs) + xlab("Percentage Poor (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 400), breaks=0:4\*100) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with many Poor have a Higher Propensity to Call",

bottom = "The Majority of Calls come from Middle Class Areas")

dev.off()

#Per in College

Percolrel <- data.frame(cbind(full.data[,17], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

Percolabs <- data.frame(cbind(full.data[,17], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerInCollege.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(Percolabs) + xlab("Percentage in College (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 600), breaks=0:6\*100) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with many College Students have a Higher Propensity to Call")

dev.off()

#Per College degs

Percolmorerel <- data.frame(cbind(full.data[,18], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

Percolmoreabs <- data.frame(cbind(full.data[,18], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerCollegePlus.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(Percolmoreabs) + xlab("Percentage with College Degree or Higher (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 500), breaks=0:5\*100) + scale\_x\_continuous(limits = c(-3, 103)),

top = "Residents in Zipcodes with many College Graduates have a Higher Propensity to Call")

dev.off()

###THIS IS BIG

#Per For Born

PerForrel <- data.frame(cbind(full.data[,19], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PerForabs <- data.frame(cbind(full.data[,19], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerForeignBorn.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(PerForabs) + xlab("Percentage in Zipcode Foreign Born (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 450), breaks=0:4\*100) + scale\_x\_continuous(limits = c(-4, 104)),

top = "Residents in Zipcodes with more Foreign Born Residents have a Higher Propensity to Call")

dev.off()

#Perc Renters

PerRentersrel <- data.frame(cbind(full.data[,22], Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PerRentersabs <- data.frame(cbind(full.data[,22], Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PerRenters.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(PerRentersabs) + xlab("Percentage in Zipcode Renting (%)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 500), breaks=0:5\*100)+ scale\_x\_continuous(limits = c(-4, 104)),

top = "Residents in Zipcodes with more Renters have a Higher Propensity to Call")

dev.off()

#Population Density

PopDensrel <- data.frame(cbind(full.data$TotPopACS/full.data[,30]/1000, Dens = full.data$count/full.data$TotPopACS)) %>% filter(complete.cases(.))

PopDensabs <- data.frame(cbind(full.data$TotPopACS/full.data[,30]/1000, Dens = full.data$count/1000)) %>% filter(complete.cases(.))

png("C:/Users/Chris/Desktop/DSO-545/Project/PopDensity.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot2(PopDensabs) + xlab("Population Density (1000x per Square Mile)") + ylab("Total Calls (Thousands)") + scale\_y\_continuous(limits = c(0, 600), breaks=0:6\*100) + scale\_x\_continuous(limits = c(-2, 55)),

top = "Residents in Zipcodes with Higher Population Densities have a Higher Propensity to Call")

dev.off()

#Overrepresented Groups

png("C:/Users/Chris/Desktop/DSO-545/Project/Overrepresentation.png", width=16, height=8, units="in", res=300, type = "cairo")

grid.arrange(

myplot3(HHIncrel) + scale\_y\_continuous(limits = c(-.05, 2.5), breaks=0:4/2) + ylab("Calls per Resident") + xlab("Median Household Income in Zipcode ($1000)"),

PerForrel %>% myplot3() + scale\_y\_continuous(limits = c(-.05, 2.5), breaks=0:4/2) + ylab("Calls per Resident") + xlab("Percentage of Foreign Born Residents in Zipcode (%)"),

top = "Calls Disproportionately Come from Areas where People are Poor and Foreign")

dev.off()