Untitled

Samantha Rodriguez

4/4/2022

head(hospitalData)

## County County.Value State.Rate Year Content.Area Date  
## 1 mohave 46 37.22 2019 Asthma 2019-04-10  
## 2 gila 46 37.22 2019 Asthma 2019-04-10  
## 3 santa cruz 45 37.22 2019 Asthma 2019-04-10  
## 4 maricopa 39 37.22 2019 Asthma 2019-04-10  
## 5 cochise 39 37.22 2019 Asthma 2019-04-10  
## 6 yuma 34 37.22 2019 Asthma 2019-04-10  
## County.Year  
## 1 mohave2019  
## 2 gila2019  
## 3 santa cruz2019  
## 4 maricopa2019  
## 5 cochise2019  
## 6 yuma2019

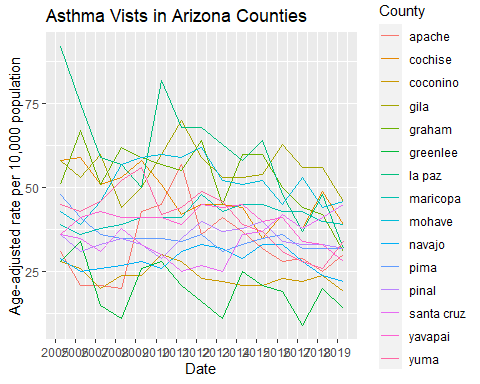
str(hospitalData)

## 'data.frame': 1535 obs. of 7 variables:  
## $ County : Factor w/ 15 levels "apache","cochise",..: 9 4 13 8 2 15 7 11 12 5 ...  
## $ County.Value: int 46 46 45 39 39 34 32 32 32 31 ...  
## $ State.Rate : num 37.2 37.2 37.2 37.2 37.2 ...  
## $ Year : chr "2019" "2019" "2019" "2019" ...  
## $ Content.Area: Factor w/ 9 levels "Asthma","Carbon Monoxide Poisoning",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Date : Date, format: "2019-04-10" "2019-04-10" ...  
## $ County.Year : chr "mohave2019" "gila2019" "santa cruz2019" "maricopa2019" ...

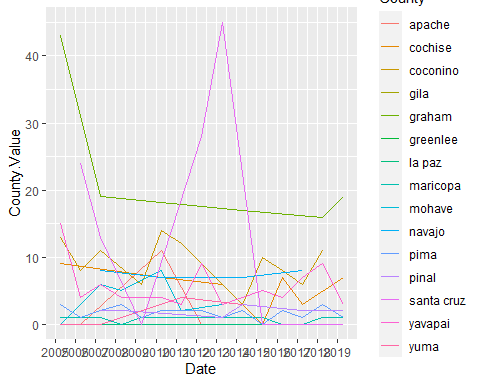
kable(str(hospitalData), format = "latex")

## 'data.frame': 1535 obs. of 7 variables:  
## $ County : Factor w/ 15 levels "apache","cochise",..: 9 4 13 8 2 15 7 11 12 5 ...  
## $ County.Value: int 46 46 45 39 39 34 32 32 32 31 ...  
## $ State.Rate : num 37.2 37.2 37.2 37.2 37.2 ...  
## $ Year : chr "2019" "2019" "2019" "2019" ...  
## $ Content.Area: Factor w/ 9 levels "Asthma","Carbon Monoxide Poisoning",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Date : Date, format: "2019-04-10" "2019-04-10" ...  
## $ County.Year : chr "mohave2019" "gila2019" "santa cruz2019" "maricopa2019" ...

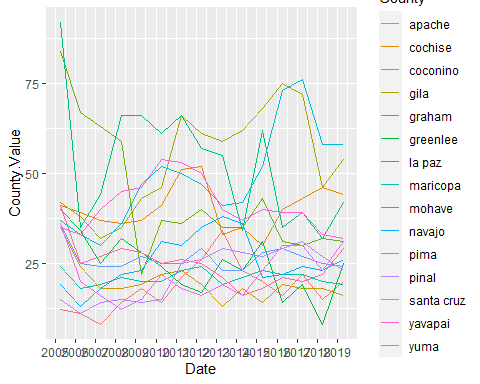
asthma <- hospitalData[hospitalData$Content.Area == "Asthma",]  
ggplot(asthma, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year") +  
 ggtitle("Asthma Vists in Arizona Counties") +  
 ylab("Age-adjusted rate per 10,000 population")



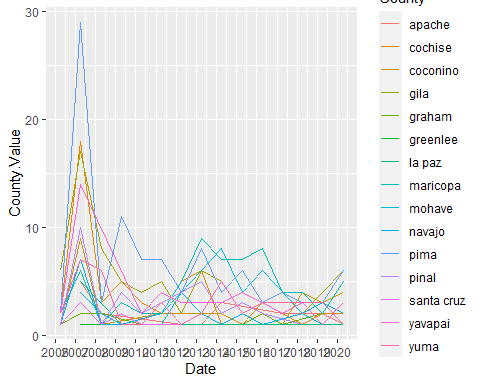
cmp <- hospitalData[hospitalData$Content.Area == "Carbon Monoxide Poisoning",]  
ggplot(cmp, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")



copd <- hospitalData[hospitalData$Content.Area == "Chronic Obstructive Pulmonary Disease (COPD)",]  
ggplot(copd, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")

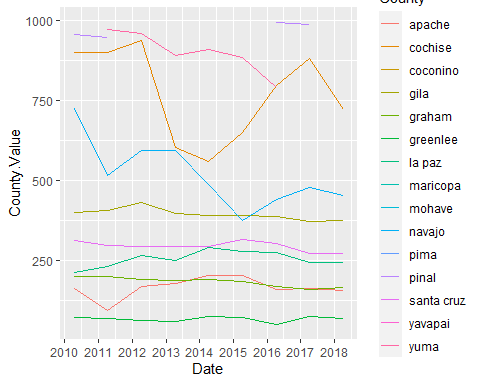


dwq <- hospitalData[hospitalData$Content.Area == "Drinking Water Quality",]  
ggplot(dwq, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")

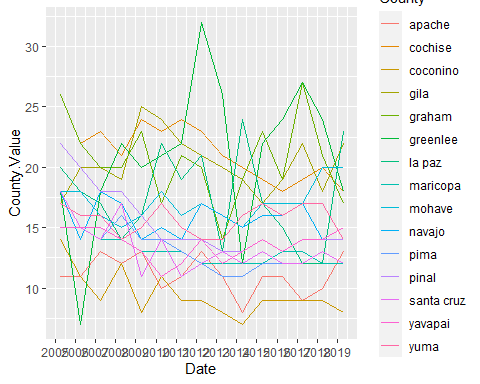


fs <- hospitalData[hospitalData$Content.Area == "Food Safety",]  
ggplot(fs, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")

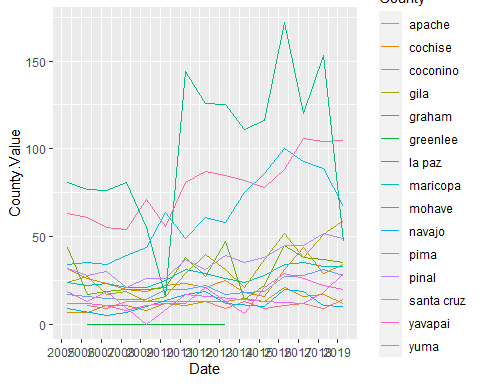
## Warning: Removed 47 row(s) containing missing values (geom\_path).



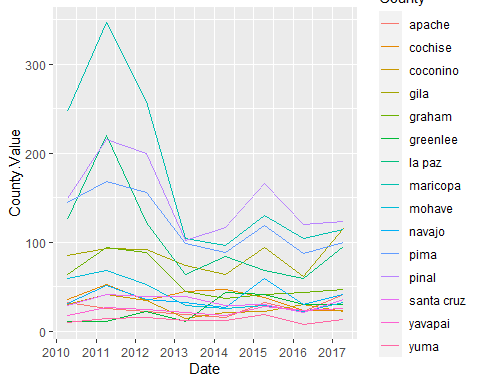
hd <- hospitalData[hospitalData$Content.Area == "Heart Disease",]  
ggplot(hd, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")



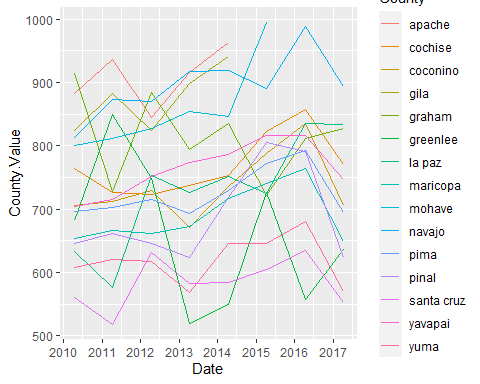
hsi <- hospitalData[hospitalData$Content.Area == "Heat Stress Illness",]  
ggplot(hsi, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")



id <- hospitalData[hospitalData$Content.Area == "Infectious Diseases",]  
ggplot(id, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")



m <- hospitalData[hospitalData$Content.Area == "Mortality",]  
ggplot(m, aes(x=Date, y=County.Value, color=County)) + geom\_line() +   
 scale\_x\_date(date\_labels = "%Y", date\_breaks = "1 year")



asthma <- hospitalData[hospitalData$Content.Area == "Asthma" &  
 hospitalData$County == "COCONINO",]  
asthma <- asthma %>% arrange(Date) %>% pull(County.Value)  
plot(asthma)  
acf(asthma)  
pacf(asthma)  
obj <- arima(asthma, order=c(1,0,0))  
acf(obj$residuals)  
predict(obj, n.ahead = 1)

hospitalNew <- hospitalData[!(hospitalData$County == "gila" |  
 hospitalData$County == "graham" |  
 hospitalData$County == "greenlee" |  
 hospitalData$County == "la paz" |  
 hospitalData$County == "santa cruz"),]  
hospitalNew <- hospitalNew[!(hospitalNew$Date == "2005-04-10" |  
 hospitalNew$Date == "2006-04-10" |  
 hospitalNew$Date == "2007-04-10" |  
 hospitalNew$Date == "2008-04-10" |  
 hospitalNew$Date == "2009-04-10" |  
 hospitalNew$Date == "2020-04-10"),]  
asthma <- hospitalNew[hospitalData$Content.Area == "Asthma",]

hospital.wider <- hospitalNew %>% select(-State.Rate)  
hospital.wider <- pivot\_wider(hospital.wider, names\_from = Content.Area,  
 values\_from = County.Value)

all.data <- full\_join(hospital.wider, censusData, by="County.Year") %>%  
 relocate(County.x, Year.x, County.y, Year.y, Date.x, Date.y)

asthma.wider <- hospital.wider %>% select(c(County.Year, Asthma))  
asthma.data <- merge(asthma.wider, censusData)  
cmp.wider <- hospital.wider %>% select(c(County.Year, `Carbon Monoxide Poisoning`))  
cmp.data <- merge(cmp.wider, censusData)  
copd.wider <- hospital.wider %>%   
 select(c(County.Year, `Chronic Obstructive Pulmonary Disease (COPD)`))  
copd.data <- merge(copd.wider, censusData)  
dwq.wider <- hospital.wider %>%   
 select(c(County.Year, `Drinking Water Quality`))  
dwq.data <- merge(dwq.wider, censusData)  
fs.wider <- hospital.wider %>%   
 select(c(County.Year, `Food Safety`))  
fs.data <- merge(fs.wider, censusData)  
hd.wider <- hospital.wider %>%   
 select(c(County.Year, `Heart Disease`))  
hd.data <- merge(hd.wider, censusData)  
hsi.wider <- hospital.wider %>%   
 select(c(County.Year, `Heat Stress Illness`))  
hsi.data <- merge(hsi.wider, censusData)  
id.wider <- hospital.wider %>%   
 select(c(County.Year, `Infectious Diseases`))  
id.data <- merge(id.wider, censusData)  
m.wider <- hospital.wider %>%   
 select(c(County.Year, `Mortality`))  
m.data <- merge(m.wider, censusData)

# colnames(cor.data) <- c('x1', 'x2', 'x3', 'x4', 'x5', 'x6', 'x7', 'x8', 'x9',  
# 'x10', 'x11', 'x12', 'x13', 'x14', 'x15', 'x16', 'x17',  
# 'x18', 'x19', 'x20', 'x21', 'x22', 'x23', 'x24',  
# 'x25', 'x26', 'x27', 'x28', 'x29', 'x30', 'x31', 'x32',  
# 'x33', 'x34', 'x35',   
# 'x36', 'x37', 'x38', 'x39', 'x40', 'x41', 'x42', 'x43',   
# 'x44', 'x45', 'x46',  
# 'x47', 'x48', 'x49', 'x50', 'x51', 'x52', 'x53', 'x54',  
# 'x55', 'x56', 'x57',  
# 'x58', 'x59', 'x60', 'x61', 'x62', 'x63', 'x64', 'x65',  
# 'x66', 'x67', 'x68',  
# 'x69', 'x70', 'x71', 'x72', 'x73', 'x74', 'x75', 'x76',  
# 'x77', 'x78', 'x79',  
# 'x80', 'x81', 'x82', 'x83', 'x84', 'x85', 'x86', 'x87',  
# 'x88', 'x89', 'x90',  
# 'x91', 'x92')

cor.data <- all.data %>% select(-c(County.Year, Year, County))  
cor.asthma <- asthma.data %>% select(-c(County.Year, Year, County))  
cor.cmp <- cmp.data %>% select(-c(County.Year, Year, County))  
cor.copd <- copd.data %>% select(-c(County.Year, Year, County))  
cor.dwq <- dwq.data %>% select(-c(County.Year, Year, County))  
cor.fs <- fs.data %>% select(-c(County.Year, Year, County))  
cor.hd <- hd.data %>% select(-c(County.Year, Year, County))  
cor.hsi <- hsi.data %>% select(-c(County.Year, Year, County))  
cor.id <- id.data %>% select(-c(County.Year, Year, County))  
cor.m <- m.data %>% select(-c(County.Year, Year, County))  
corrplot::corrplot(cor(cor.asthma, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.cmp, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.copd, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.dwq, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.fs, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.hd, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.hsi, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.id, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.m, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
corrplot::corrplot(cor(cor.data, use = 'pairwise.complete.obs'), tl.cex = 0.8)  
cor.v <- cor(cor.data)  
pos.cor <- cor.v[cor.v > 0.5]

test <- cor.test(hospital.wider$Asthma, censusData$Total.Population)  
test$p.value  
test$estimate  
cor.test(hospital.wider[,4], censusData$Total.Population)

all.data$Asthma <- as.numeric(all.data$Asthma)  
all.data <- all.data %>% select(-c(County.x, Year.x, Date.x, `Food Safety`))  
colnames(all.data)[1] <- "County"  
colnames(all.data)[2] <- "Year"  
colnames(all.data)[3] <- "Date"  
all.data$Year <- as.factor(all.data$Year)  
names(all.data) <- gsub(" ", ".", names(all.data))  
colnames(all.data)[7] <- "COPD"

correlation <- data.frame(Content.Area = character(),  
 Demographic = character(),  
 p.value = integer(),  
 corr = integer())  
for(cont in 5:12){  
 for(demo in 13:95){  
 #print(paste(cont, demo))  
 hold <- cor.test(pull(all.data[,cont]), pull(all.data[,demo]))  
 corr.df <- data.frame(Content.Area = c(colnames(all.data)[cont]),  
 Demographic = c(colnames(all.data)[demo]),  
 p.value = c(hold$p.value),  
 corr = c(hold$estimate))  
 if(is.na(correlation[1,1])){  
 correlation <- corr.df  
 }  
 else {  
 correlation <- rbind(correlation, corr.df)  
 }  
   
 }  
}

correlation$adj.p.value <- p.adjust(correlation$p.value, method="bonferroni")

adj.cor <- correlation %>% filter(adj.p.value < 0.05) %>% arrange(corr)

#ggplot(all.data, aes(x=County, y=Asthma)) + geom\_boxplot()  
for (index in 5:12) {  
 print(ggplot(all.data, aes\_string(x="County", y=colnames(all.data)[index])) +  
 geom\_boxplot())  
}

obj <- lm(Asthma ~ County, data=all.data)  
summary(obj)  
plot(obj)  
anova(obj)  
emmeans(obj, pairwise ~ County)

content.area.p <- data.frame(  
 County = factor(),  
 p.value = integer()  
)  
for (index in 5:12) {  
 obj <- lm(paste(colnames(all.data)[index], "~", 'County'), data=all.data)  
 sum.obj <- summary(obj)  
 f <- sum.obj$fstatistic  
 p <- pf(f[1], f[2], f[3], lower.tail=F)  
 p.df <- data.frame(County = colnames(all.data)[index],  
 p.value = p)  
 if(index == 5)  
 {  
 content.area.p <- p.df  
 }  
 else  
 {  
 content.area.p <- rbind(content.area.p, p.df)  
 }  
 print(sum.obj)  
 print(anova(obj))  
 print(plot(obj))  
 #print(emmeans(obj, pairwise ~ County))  
}

content.area.p <- content.area.p %>% arrange(p.value)

obj <- lm(Asthma ~ County, data=all.data)  
letter.data.asthma <- emmeans(obj, specs = ~ County) %>%  
 multcomp::cld(Letters = letters, level = 0.95) %>%  
 mutate(.group = str\_remove\_all(.group, '\\s')) %>%  
 mutate(y=73)

ggplot(all.data, aes(x=reorder(County, Asthma), y=Asthma)) +  
 geom\_boxplot() +  
 labs(x = 'County', y = 'Asthma') +  
 geom\_text(data=letter.data.asthma, aes(x=County, y=y, label=.group), size=8)

for(iter in 5:12)  
{  
 obj <- lm(paste(colnames(all.data)[iter], "~", 'County'), data=all.data)  
 letter.data <- emmeans(obj, specs = ~ County) %>%  
 multcomp::cld(Letters = letters, level = 0.95) %>%  
 mutate(.group = str\_remove\_all(.group, '\\s')) %>%  
 mutate(y = max(all.data[,iter], na.rm=T) + (max(all.data[,iter], na.rm=T) \* .05))  
 hold.data <- data.frame(  
 County = all.data$County,  
 Value = all.data[,iter]  
 )  
 hold.data <- hold.data %>% na.omit()  
 print(ggplot(hold.data,   
 aes(x=reorder(County, !!sym(colnames(hold.data)[2]), FUN = median),   
 y=!!sym(colnames(hold.data)[2]))) +  
 geom\_boxplot() +  
 labs(x = "County") +  
 geom\_text(data = letter.data, aes(x=County, y=y, label=.group),   
 size = 8))  
}

final.p.value <- adj.cor %>%   
 group\_by(Content.Area) %>% slice(which.min(adj.p.value)) %>%  
 arrange(adj.p.value)

ggplot(all.data, aes(x=COPD, y=Median.Age, color=County)) + geom\_point()  
ggplot(all.data, aes(x=Infectious.Diseases, y=Under.5.Population, color=County)) +  
 geom\_point()  
ggplot(all.data, aes(x=Asthma, y=Median.Age, color=County)) + geom\_point()

for (index in 1:7) {  
 print(ggplot(all.data,   
 aes\_string(x=pull(final.p.value[index,1]),   
 y=pull(final.p.value[index,2]))) +  
 geom\_point() + geom\_smooth(method="lm", model='y~x'))   
}

ggplot(all.data,   
 aes(x=Heat.Stress.Illness, y=AI.and.AN)) +  
 geom\_point() + geom\_smooth(method="loess", model='y~x')

hospital.info <- str(hospital.wider)  
write.table(hospital.info, file = "hospital.txt", sep = ",", quote = FALSE, row.names = T)