visualizations

Sarah Sandmaier

2025-03-28

library(ggplot2)  
library(dplyr)  
library(ggtext)  
library(tidyr)  
library(flextable)  
library(officer)

saving and uploading files

VGT <- read.csv("VGT\_stock\_data\_cleaned.csv")  
VHT <- read.csv("VHT\_stock\_data\_cleaned.csv")  
ICSA <- read.csv("ICSA\_Initial\_claims\_cleaned.csv")  
Unrate <- read.csv("unemployment\_rate\_data.csv")  
CovState <- read.csv("Weekly\_United\_States\_COVID-19\_Cases\_and\_Deaths\_by\_State\_-\_ARCHIVED\_20250304\_cleaned.csv")  
CovGroup <- read.csv("COVID-19\_Weekly\_Cases\_and\_Deaths\_by\_Age\_\_Race\_Ethnicity\_\_and\_Sex\_-\_ARCHIVED\_20250310\_cleaned.csv"")

Graph of VGT vs VHT

vgt\_vht <- ggplot() +   
 geom\_rect(aes(xmin = as.Date("2020-03-01"), xmax = as.Date("2023-12-01"), ymin = -Inf, ymax = Inf),   
 fill = "red", alpha = 0.3) +  
 geom\_text(aes(x = as.Date("2022-01-01"), y = max(VHT$average\_movement, na.rm = TRUE),   
 label = "COVID-19", color = "black"),   
 size = 4, fontface = "bold", vjust = -15) +   
 geom\_line(data = VHT, aes(x = Date, y = average\_movement, color = "VHT", group = 1), linewidth = 1) +   
 geom\_line(data = VGT, aes(x = Date, y = average\_movement, color = "VGT", group = 1), linewidth = 1) +   
 labs(title = "Average movement of Stock Markets",  
 x = "Date",  
 y = "Price Average Movement",  
 color = "Stock") +   
 scale\_color\_manual(values = c("VHT" = "blue", "VGT" = "green")) +   
 theme\_minimal() +  
 theme(  
 panel.grid.major = element\_blank(),   
 panel.grid.minor = element\_blank(),   
 axis.line = element\_line(color = "black"),   
 legend.position = "right",   
 plot.title = element\_text(hjust = 0.5, size = 14, face = "bold"),   
 plot.caption = element\_markdown(hjust = 0.5, size = 14, face = "plain"),  
 plot.caption.position = "plot",   
 plot.margin = margin(t = 10, r = 10, b = 50, l = 10)   
 )

Sex comparison graphs

sexg <- ggplot(CovGroup, aes(x = sex, y = case\_count\_suppressed, fill= sex)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Sex", y = "Case Count (per 100k)", x = NULL) +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +  
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
sexd <- ggplot(CovGroup, aes(x = sex, y = death\_count\_suppressed, fill= sex)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Sex", y = "Death Count (per 100k)", x = NULL) +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
  
grid.arrange(  
 sexg + theme(legend.position = "none") + labs(title = NULL),  
 sexd + labs(title = NULL),  
 ncol = 2,  
 bottom = textGrob("Covid-19 Cases by Sex", gp = gpar(fontsize = 14, fontface = "bold"))  
)

race comparison graphs

raceg <- ggplot(CovGroup, aes(x = race\_ethnicity\_combined, y = case\_count\_suppressed, fill= race\_ethnicity\_combined)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Race/Ethnicity", y = "Case Count (per 100k)", x = NULL, fill = "Race/Ethnicity") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
raced <- ggplot(CovGroup, aes(x = race\_ethnicity\_combined, y = death\_count\_suppressed, fill= race\_ethnicity\_combined)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Race/Ethnicity", y = "Death Count (per 100k)", x = NULL, fill = "Race/Ethnicity") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5,face = "bold"))  
  
grid.arrange(  
 raceg + theme(legend.position = "none") + labs(title = NULL),  
 raced + labs(title = NULL),  
 ncol = 2,  
 bottom = textGrob("Covid-19 Cases by Race/Ethnicity", gp = gpar(fontsize = 14, fontface = "bold"))  
)

age comparison graphs

custom\_order <- c("0 - 4 Years","5 - 11 Years","12 - 15 Years","16 - 17 Years",  
 "18 - 29 Years", "30 - 39 Years","40 - 49 Years","50 - 64 Years",  
 "65 - 74 Years", "75+ Years")  
  
CovGroup <- CovGroup %>%  
 mutate(age\_group = factor(age\_group, levels = custom\_order))  
  
  
ageg <- ggplot(CovGroup, aes(x = age\_group, y = case\_count\_suppressed, fill= age\_group)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Age Group", y = "Case Count (per 100k)", x = NULL, fill = "Age Group") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
aged <- ggplot(CovGroup, aes(x = age\_group, y = death\_count\_suppressed, fill= age\_group)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Age Group", y = "Death Count (per 100k)", x = NULL, fill = "Age Group") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
grid.arrange(  
 ageg + theme(legend.position = "none") + labs(title = NULL),  
 aged + theme(legend.position = "bottom") + labs(title = NULL),  
 nrow = 2,  
 bottom = textGrob("Covid-19 Cases by Age Group", gp = gpar(fontsize = 14, fontface = "bold"))  
)

jurisdiction comparison graphs

reg\_order <- c("Region 1","Region 2","Region 3","Region 4","Region 5","Region 6",  
 "Region 7","Region 8","Region 9","Region 10")  
  
CovGroup <- CovGroup %>%  
 mutate(jurisdiction = factor(jurisdiction, levels = reg\_order))  
  
regg <- ggplot(CovGroup, aes(x = jurisdiction, y = case\_count\_suppressed, fill= jurisdiction)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Region Group", y = "Case Count (per 100k)", x = NULL, fill = "jurisdiction") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
regd <-ggplot(CovGroup, aes(x = jurisdiction, y = death\_count\_suppressed, fill= jurisdiction)) +  
 geom\_bar(stat = "identity") +   
 labs(title = "Region ", y = "Death Count (per 100k)", x = NULL, fill = "jurisdiction") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-5)) +   
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))  
  
grid.arrange(  
 regg + theme(legend.position = "none") + labs(title = NULL),  
 regd + theme(legend.position = "bottom") + labs(title = NULL),  
 nrow = 2,  
 bottom = textGrob("Covid-19 Cases by Region", gp = gpar(fontsize = 14, fontface = "bold"))  
)

ICSA claims vs unrate graphs

unem\_bar <- ggplot() +  
 geom\_rect(aes(xmin = as.Date("2020-03-01"), xmax = as.Date("2023-12-01"), ymin = -Inf, ymax = Inf),   
 fill = "red", alpha = 0.15) +  
 geom\_text(aes(x = as.Date("2022-01-01"),   
 y = max(ICSA$icsa, na.rm = TRUE)),   
 label = "COVID-19", color = "black",   
 size = 4, fontface = "bold", vjust = 1) +   
 geom\_col(data = ICSA, aes(x = observation\_date, y = icsa), fill = "blue") +   
 labs(title = "ICSA Unemployment Claims", x = "Date", y = "Case Amount (in 10k)") +  
 scale\_y\_continuous(labels = label\_number(scale = 1e-4)) +  
 theme\_minimal()  
  
  
unrate\_line <- ggplot() +  
 geom\_rect(aes(xmin = as.Date("2020-03-01"), xmax = as.Date("2023-12-01"), ymin = -Inf, ymax = Inf),   
 fill = "red", alpha = 0.15) +  
 geom\_text(aes(x = as.Date("2022-01-01"),   
 y = max(Unrate$UNRATE, na.rm = TRUE)),   
 label = "COVID-19", color = "black",   
 size = 4, fontface = "bold", vjust = 1) +   
 geom\_line(data = Unrate, aes(x = observation\_date, y = UNRATE), color = "purple") +   
 labs(title = "Unemployment Rate", x = "Date", y = "Unemployment Rate (%)") +  
 theme\_minimal()  
  
grid.arrange(  
 unem\_bar + theme(legend.position = "none") + labs(title = NULL),  
 unrate\_line + labs(title = NULL),  
 nrow = 1,  
 bottom = textGrob("Unemployment Claims vs Rate", gp = gpar(fontsize = 14, fontface = "bold"))  
)

region to state table

region\_list <- list(  
 "Region 1" = c("ME", "MA", "RI", "NH", "PR"),  
 "Region 2" = c("VT", "NY", "CT"),  
 "Region 3" = c("NJ", "DE", "PA", "VI"),  
 "Region 4" = c("MD", "NC", "SC", "VA", "WV", "AL", "FL", "GA"),  
 "Region 5" = c("LA", "MS", "TX"),  
 "Region 6" = c("KY", "MI", "OH", "TN"),  
 "Region 7" = c("IL", "IN", "WI"),  
 "Region 8" = c("AR", "IA", "MN", "MO", "NE", "ND", "SD"),  
 "Region 9" = c("AK", "AZ", "CA", "GU", "HI", "ID", "MT", "NV", "OR", "WA"),  
 "Region 10" = c("CO", "KS", "NM", "OK", "UT", "WY")  
)  
  
region\_df <- region\_list %>%  
 tibble::enframe(name = "Region", value = "State") %>%  
 unnest(cols = c(State))  
  
region\_wide <- region\_df %>%  
 group\_by(Region) %>%  
 mutate(row = row\_number()) %>%  
 pivot\_wider(names\_from = Region, values\_from = State) %>%  
 select(-row)  
region\_table <- flextable(region\_wide) %>%  
 theme\_vanilla() %>%   
 set\_caption("States by Region") %>%   
 border\_remove() %>%   
 border(i = 1, border.top = fp\_border(width = 1)) %>%   
 bg(j = seq(1, ncol(region\_wide), by = 2), bg = "#e0f7fa") %>%   
 bg(j = seq(1, ncol(region\_wide), by = 2), part = "header", bg = "#e0f7fa") %>%   
 autofit() %>%  
 align(align = "center", part = "all")