Data Translation Project - Group 6

## Research Question 1

How has COVID affected the health of the retail industry, as measured by employment?

### Data Cleaning

To examine the impact of COVID-19 on retail employment, we utilized data from the IPUMS Current Population Survey. The data was filtered to include individuals aged 18-62 who were in the labor force and worked in the retail industry (NAICS codes 4670-5790). Key variables were created, including a PostCOVID indicator for observations from March 2020 onward, age groups, gender, and total employment status.

if (!require("ipumsr")) stop("Reading IPUMS data into R requires the ipumsr package. It can be installed using the following command: install.packages('ipumsr')")

Loading required package: ipumsr

Warning: package 'ipumsr' was built under R version 4.2.3

library(rio)  
library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

library(tidyverse)

── Attaching packages  
───────────────────────────────────────  
tidyverse 1.3.2 ──

✔ ggplot2 3.4.0 ✔ purrr 1.0.1  
✔ tibble 3.1.8 ✔ stringr 1.5.0  
✔ tidyr 1.2.1 ✔ forcats 0.5.2  
✔ readr 2.1.3   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()

library(vtable)

Loading required package: kableExtra  
  
Attaching package: 'kableExtra'  
  
The following object is masked from 'package:dplyr':  
  
 group\_rows

library(fixest)

Warning: package 'fixest' was built under R version 4.2.3

library(RColorBrewer)  
library(marginaleffects)

Warning: package 'marginaleffects' was built under R version 4.2.3

library(lubridate)

Loading required package: timechange  
  
Attaching package: 'lubridate'  
  
The following objects are masked from 'package:base':  
  
 date, intersect, setdiff, union

library(sandwich)  
library(lmtest)

Loading required package: zoo  
  
Attaching package: 'zoo'  
  
The following objects are masked from 'package:base':  
  
 as.Date, as.Date.numeric

library(patchwork)  
  
ddi <- read\_ipums\_ddi("cps\_00002.xml")  
data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

ind <- read\_csv("indnames.csv")

Rows: 8185 Columns: 2  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (1): indname  
dbl (1): ind  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

industry\_df <- import('indnames.csv')  
  
df\_final<- data %>% inner\_join( industry\_df, by = c("IND"="ind")) %>%   
 filter(YEAR >= 2018,!is.na(INCTOT))

# Data Cleaning  
retail\_data <- data %>%  
 filter(IND >= 4670 & IND <= 5790, AGE >= 18 & AGE <= 62, LABFORCE == 2) %>%  
 mutate(PostCOVID = ifelse(YEAR >= 2020 & MONTH >= 3, 1, 0),  
 AgeGroup = case\_when(AGE >= 18 & AGE <= 40 ~ 0, AGE >= 41 & AGE <= 62 ~ 1),  
 Gender = case\_when(SEX == 1 ~ 1, SEX == 2 ~ 0,TRUE ~ NA\_real\_),  
 TotalEmployed= ifelse(EMPSTAT %in% c(10,12),1,0),  
 YearMonth = as.yearmon(paste(YEAR, sprintf("%02d", MONTH), "01", sep = "-")))

The data was then aggregated to the year-month level, with employment rates calculated as the weighted mean of the total employed using the WTFINL survey weights. This ensures the employment measures are representative of the population.

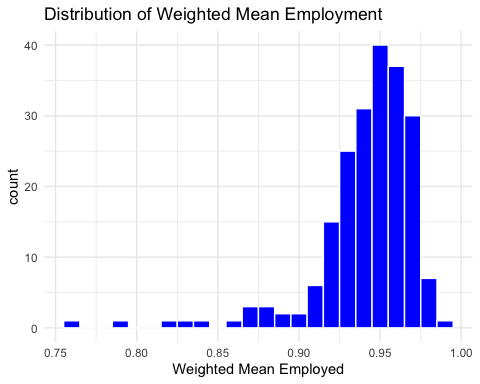
# Regression model  
regression <- retail\_data %>%   
 filter(!is.na(WTFINL),!is.na(TotalEmployed)) %>%   
 group\_by(YEAR, YearMonth, AgeGroup, Gender, PostCOVID) %>%  
 summarize(wtemployed = weighted.mean(TotalEmployed, WTFINL, na.rm = TRUE))

`summarise()` has grouped output by 'YEAR', 'YearMonth', 'AgeGroup', 'Gender'.  
You can override using the `.groups` argument.

### Distribution of the weighted employment rate

Examining the distribution of the weighted employment rate reveals a left-skewed pattern, with a concentration of values around 0.91-0.98 and a longer left tail. This suggests that while employment rates are generally high, there are some notably lower values, likely corresponding to the pandemic period. The histogram’s bins show a peak frequency around 0.94-0.96, with employment rates becoming progressively less common moving further left. This skewed shape indicates that the pandemic’s impact has pulled down the typical employment rate.

# Distribution of weighted employed   
ggplot(regression, aes(x = wtemployed)) +  
 geom\_histogram(binwidth = 0.01, fill = "blue", color = "white") +  
 labs(title = "Distribution of Weighted Mean Employment", x = "Weighted Mean Employed") +  
 theme\_minimal()



### Overall Employment Trend Plots

To investigate the impact of the COVID-19 pandemic on the health of the retail industry, we created three graphs that provide a comprehensive view of the employment trend. The first two graphs visualize the monthly trends of employment and unemployment ratios, while the third graph examines the effect of the pandemic on the retail workforce by gender.

# Analyzing monthly trends in employment and unemployment, grouped by year-month  
yearly\_data <- retail\_data %>%  
 group\_by(YEAR, YearMonth, PostCOVID) %>%  
 summarize(TotalLaborForce = sum(LABFORCE %in% c(1, 2)),  
 TotalEmployed = sum(EMPSTAT %in% c(10, 12)),  
 TotalUnemployed = sum(EMPSTAT %in% c(20, 21)),  
 .groups = 'drop')  
  
# Monthly Trend of Employment Ratio Graph  
employment\_plot <- ggplot(yearly\_data, aes(x = YearMonth, y = yearly\_data$TotalEmployed / yearly\_data$TotalLaborForce, group = 1)) +  
 geom\_line(color = "darkorange", size = 1) +  
 geom\_point(color = "steelblue2", size = 1) +  
 geom\_vline(xintercept = as.numeric(as.yearmon("Mar 2020")), linetype = "dashed", color = "red") +  
 annotate("text", x = as.yearmon("Mar 2020"), y = 0.9, label = "March 2020", color = "darkgreen", angle = 90, vjust = -0.5, hjust = 0) +  
 labs(title = "Monthly Trend of Employment Ratio", x = "Year-Month", y = "Employment Ratio") +  
 theme\_classic()

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
ℹ Please use `linewidth` instead.

# Monthly Trend of Unemployment Ratio Graph  
unemployment\_plot <- ggplot(yearly\_data, aes(x = YearMonth, y = yearly\_data$TotalUnemployed / yearly\_data$TotalLaborForce, group = 1)) +  
 geom\_line(color = "darkorange", size = 1) +  
 geom\_point(color = "steelblue2", size = 1) +  
 geom\_vline(xintercept = as.numeric(as.yearmon("Mar 2020")), linetype = "dashed", color = "red") +  
 annotate("text", x = as.yearmon("Mar 2020"), y = 0.1, label = "March 2020", color = "darkgreen", angle = 90, vjust = -0.5, hjust = 0) +  
 labs(title = "Monthly Trend of Unemployment Ratio", x = "Year-Month", y = "Unemployment Ratio") +  
 theme\_classic()  
  
# Combining the 2 graphs  
employment\_plot + unemployment\_plot + plot\_layout(ncol = 2)

Warning: Use of `yearly\_data$TotalEmployed` is discouraged.  
ℹ Use `TotalEmployed` instead.

Warning: Use of `yearly\_data$TotalLaborForce` is discouraged.  
ℹ Use `TotalLaborForce` instead.

Warning: Use of `yearly\_data$TotalEmployed` is discouraged.  
ℹ Use `TotalEmployed` instead.

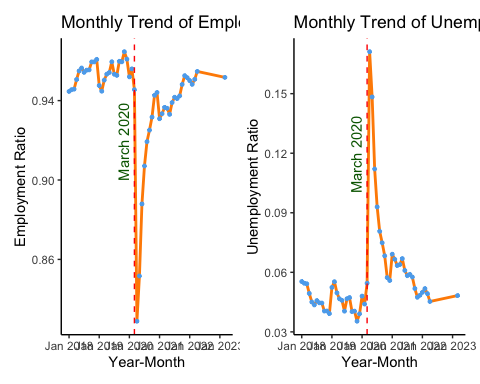
Warning: Use of `yearly\_data$TotalLaborForce` is discouraged.  
ℹ Use `TotalLaborForce` instead.

Warning: Use of `yearly\_data$TotalUnemployed` is discouraged.  
ℹ Use `TotalUnemployed` instead.

Warning: Use of `yearly\_data$TotalLaborForce` is discouraged.  
ℹ Use `TotalLaborForce` instead.

Warning: Use of `yearly\_data$TotalUnemployed` is discouraged.  
ℹ Use `TotalUnemployed` instead.

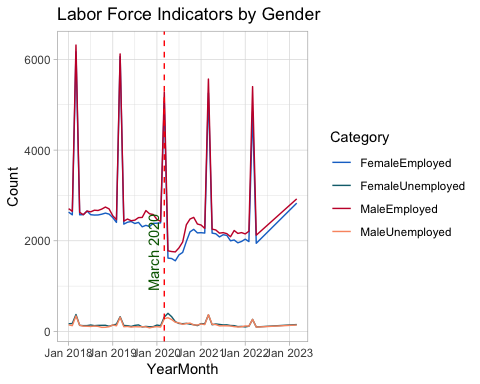
Warning: Use of `yearly\_data$TotalLaborForce` is discouraged.  
ℹ Use `TotalLaborForce` instead.



# Exploring gender-specific dynamics in the labor market  
gender\_specific <- retail\_data %>%  
 group\_by(YEAR, YearMonth, PostCOVID) %>%  
 summarize(  
 TotalLaborForce = sum(LABFORCE %in% c(1, 2)),  
 TotalEmployed = sum(EMPSTAT %in% c(10, 12)),  
 TotalUnemployed = sum(EMPSTAT %in% c(20, 21)),  
 MaleEmployed = sum(SEX == 1 & EMPSTAT %in% c(10, 12)),  
 FemaleEmployed = sum(SEX == 2 & EMPSTAT %in% c(10, 12)),  
 MaleUnemployed = sum(SEX == 1 & EMPSTAT %in% c(20, 21)),  
 FemaleUnemployed = sum(SEX == 2 & EMPSTAT %in% c(20, 21))) %>%  
 ungroup()

`summarise()` has grouped output by 'YEAR', 'YearMonth'. You can override using  
the `.groups` argument.

# Make gender variables into long format  
gender\_long <- gender\_specific %>%   
 pivot\_longer(c(MaleEmployed, FemaleEmployed, MaleUnemployed, FemaleUnemployed),  
 names\_to = "category", values\_to = "Count")   
  
# Labor Force Indicators by Gender Graph   
ggplot(gender\_long, aes(x = YearMonth, y = Count, color = category)) +   
 geom\_line() +  
 scale\_color\_manual(values = c("#1874cd", "#106e7c","#c71936", "#fa9770")) +  
 geom\_vline(xintercept = as.numeric(as.yearmon("Mar 2020")), linetype = "dashed", color = "red") +  
 annotate("text", x = as.yearmon("Mar 2020"), y = 900, label = "March 2020", color = "darkgreen", angle = 90, vjust = -0.5, hjust = 0) +  
 labs(title = "Labor Force Indicators by Gender", y = "Count", color = "Category") +   
 theme\_light()



The first two graphs depict the monthly trends of employment and unemployment ratios in the retail industry, respectively. The employment ratio graph shows a relatively stable trend before the pandemic, fluctuating around 0.94 - 0.96. However, in March 2020, there is a sharp decline in the employment ratio, dropping to approximately 0.82 by April 2020. This sudden decrease aligns with the onset of the COVID-19 pandemic and the implementation of lockdown measures, which significantly impacted the retail sector. Similarly, the unemployment ratio graph shows a consistently low ratio between 0.04 - 0.06 prior to the pandemic, followed by a spike reaching over 0.18 in April 2020.

Following the initial shock, both graphs show a gradual recovery in employment. The employment ratio begins to trend upward, while the unemployment ratio starts to decline. This suggests that as businesses adapted to the new circumstances and lockdown measures were eased, the retail industry started to regain some of the lost jobs. However, it is important to note that the employment ratio does not immediately return to pre-pandemic levels, indicating that the impact of COVID-19 on retail employment persisted even as the situation improved.

The third graph breaks down the labor force indicators by gender, providing more insights on how the pandemic affected employment and unemployment for males and females in the retail sector. Before the pandemic, the counts of employed males and females remained relatively stable, with males slightly exceeding females in retail workforce. This indicates that men make up a larger portion of the retail workforce. However, the graph reveals that the impact of the pandemic on employment was higher for females compared to males. The decline in female employment during the pandemic period is steeper and more sustained than the decline in male employment, suggesting that women in the retail industry were disproportionately affected by job losses. Additionally, the recovery in employment counts appears to be slower for females, indicating that women faced greater challenges in regaining employment during the recovery phase.

### Regression Models

In this regression analysis, we aim to investigate the impact of the COVID-19 pandemic on the weighted employment (wtemployed) in the retail industry while considering the effects of age group and gender. The analysis includes three models with different specifications.

Controlling for YearMonth: We control for the YearMonth variable in all three models to account for the temporal effects on employment. By including YearMonth, we capture the overall trend in employment over time, which helps isolate the impact of the COVID-19 pandemic (PostCOVID) and other variables of interest.

Interaction Terms: In models 2 and 3, we introduce interaction terms between PostCOVID and AgeGroup, as well as PostCOVID and Gender. These interaction terms allow us to examine whether the impact of the COVID-19 pandemic on employment varies across different age groups and genders. Since there is only one retail industry in the data, including interaction terms while controlling for YearMonth enables us to capture the differential effects of the pandemic on specific subgroups within the retail industry.

Log Transformation: In model 3, we apply a log transformation to the dependent variable wtemployed. This transformation is useful when dealing with skewed data, as it helps to normalize the distribution and reduce the influence of extreme values. Using log(wtemployed) as the dependent variable enables us to interpret the coefficients as percentage changes in employment for a unit change in the independent variables.

# Regression models  
model1 <- feols(wtemployed ~ PostCOVID + YearMonth, vcov = "hetero" ,data = regression)  
model2 <- feols(wtemployed ~ PostCOVID\*AgeGroup + PostCOVID\*Gender + YearMonth, vcov = "hetero", data = regression)  
model3 <- feols(log(wtemployed) ~ PostCOVID\*AgeGroup + PostCOVID\*Gender + YearMonth, vcov = "hetero", data = regression)  
summary(model1)

OLS estimation, Dep. Var.: wtemployed  
Observations: 208   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -9.889246 4.784824 -2.06679 4.0010e-02 \*   
PostCOVID -0.037499 0.007541 -4.97287 1.3921e-06 \*\*\*  
YearMonth 0.005369 0.002370 2.26574 2.4512e-02 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.027818 Adj. R2: 0.209057

summary(model2)

OLS estimation, Dep. Var.: wtemployed  
Observations: 208   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -9.904434 4.363270 -2.269957 2.4271e-02 \*   
PostCOVID -0.043840 0.010371 -4.226946 3.5935e-05 \*\*\*  
AgeGroup 0.023110 0.002481 9.314108 < 2.2e-16 \*\*\*  
Gender 0.007265 0.002481 2.928058 3.8043e-03 \*\*   
YearMonth 0.005369 0.002161 2.484597 1.3787e-02 \*   
PostCOVID:AgeGroup 0.008291 0.007625 1.087340 2.7819e-01   
PostCOVID:Gender 0.004390 0.007625 0.575709 5.6546e-01   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.023886 Adj. R2: 0.40525

summary(model3)

OLS estimation, Dep. Var.: log(wtemployed)  
Observations: 208   
Standard-errors: Heteroskedasticity-robust   
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -12.386965 4.874121 -2.541374 1.1796e-02 \*   
PostCOVID -0.049233 0.011982 -4.108991 5.7841e-05 \*\*\*  
AgeGroup 0.024283 0.002680 9.061079 < 2.2e-16 \*\*\*  
Gender 0.007615 0.002680 2.841365 4.9549e-03 \*\*   
YearMonth 0.006102 0.002414 2.527906 1.2244e-02 \*   
PostCOVID:AgeGroup 0.010112 0.008616 1.173605 2.4194e-01   
PostCOVID:Gender 0.005524 0.008616 0.641117 5.2218e-01   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 0.026894 Adj. R2: 0.388581

etable(model1, model2, model3)

model1 model2  
Dependent Var.: wtemployed wtemployed  
   
Constant -9.889\* (4.785) -9.904\* (4.363)  
PostCOVID -0.0375\*\*\* (0.0075) -0.0438\*\*\* (0.0104)  
YearMonth 0.0054\* (0.0024) 0.0054\* (0.0022)  
AgeGroup 0.0231\*\*\* (0.0025)  
Gender 0.0073\*\* (0.0025)  
PostCOVID x AgeGroup 0.0083 (0.0076)  
PostCOVID x Gender 0.0044 (0.0076)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob. Heteroskedast.-rob.  
Observations 208 208  
R2 0.21670 0.42249  
Adj. R2 0.20906 0.40525  
  
 model3  
Dependent Var.: log(wtemployed)  
   
Constant -12.39\* (4.874)  
PostCOVID -0.0492\*\*\* (0.0120)  
YearMonth 0.0061\* (0.0024)  
AgeGroup 0.0243\*\*\* (0.0027)  
Gender 0.0076\*\* (0.0027)  
PostCOVID x AgeGroup 0.0101 (0.0086)  
PostCOVID x Gender 0.0055 (0.0086)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 208  
R2 0.40630  
Adj. R2 0.38858  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Regression Models Interpretation

Model 1: The baseline model estimates a 0.038 reduction in the weighted employment rate after the start of COVID, which is statistically significant at the 0.1% level. This model also controls for the overall time trend, with each additional month associated with a 0.005 increase in the weighted employment rate, statistically significant at the 5% level.

Model 2: This model extends the baseline model by including interaction terms between the PostCOVID indicator and the AgeGroup and Gender variables, allowing the effect of COVID to vary by age group and gender. The model also controls for the overall time trend. The main effect of PostCOVID remains significant, indicating a 0.044 decrease in the weighted employment rate after the start of COVID, holding age group, gender, and time trend constant. The main effects of AgeGroup and Gender are also significant, with each additional age group associated with a 0.024 higher employment rate and being male associated with a 0.007 higher employment rate, relative to being female. However, the interaction terms are not statistically significant, suggesting that the effect of COVID on employment does not significantly differ by age group or gender. The overall time trend remains significant, with each additional month associated with a 0.005 increase in the weighted employment rate.

Model 3: Given the skewed distribution of the weighted employment rate, this model uses a log-transformed dependent variable. The log-linear specification estimates that the post-COVID period saw an overall 4.9% decline in retail employment, statistically significant at the 0.1% level, holding age group, gender, and the time trend constant. Consistent with the other models, it also finds higher employment rates for each additional age group (by 2.5%, statistically significant at the 0.1% level) and for males (by 0.7%, statistically significant at the 1% level). The interaction terms remain insignificant, indicating that the pandemic’s effect is relatively uniform across age groups and genders.

### Limitations

When interpreting the results of this analysis, it is important to consider potential biases and limitations. For instance, the pre-post comparison assumes that the pandemic was the only major factor differentially affecting retail employment around March 2020, but there could be other time-varying confounders not captured in the data. In addition, the models account for age and gender but do not capture other potentially relevant characteristics such as race, education, or geographic location. Consequently, the parametric assumptions of the models may not fully capture the complexity of the true relationships. Lastly, sample selection bias is a potential concern in this analysis, as the data may not be perfectly representative of the entire retail industry population. If the sample overrepresents certain retail sectors or geographic regions that were more severely affected by the pandemic, the estimated impact of COVID-19 on employment may be overstated.

### Conclusion

Across all models, the PostCOVID coefficient remains negative and statistically significant at the 0.1% level, indicating a substantial decline in retail employment after the start of the pandemic. The magnitude of the effect ranges from a 0.038 reduction in the weighted employment rate in model 1 to a 4.9% decrease in model 3. The time trend (YearMonth) is positive and significant at the 5% level in all models, suggesting a gradual increase in employment over time, holding other factors constant. Models 2 and 3 introduce age group and gender as control variables, revealing that older age groups (41 - 62) and males have significantly higher employment rates compared to their respective counterparts. However, the interaction terms between PostCOVID and AgeGroup/Gender are not statistically significant, indicating that the pandemic’s impact on employment is relatively uniform across these demographic groups. The use of heteroskedasticity-robust standard errors in all models strengthens the reliability of these findings by accounting for potential non-constant variance in the errors.

## Research Question 2: How has retail fared relative to other industries?

### Analysis Approach

Based on our interpretation of the research question, we believe a logit interaction model with a binary dependent variable would be able to effectively answer our question. Similar to the first question, we will be using employment within industries to help measure the vitality of the retail industry. In this model however, we will be creating a new variable called ‘employed’, which is true or false depending on a person’s employment status (EMPSTAT). Our dependent variables would consist of binary indicators: “Retail,” representing whether a person is or was employed in the retail industry, and “COVID,” indicating whether the data observation was recorded before or during the COVID-19 pandemic. By examining the interaction between “Retail” and “COVID,” we can assess the impact of the COVID-19 pandemic on employment within the retail sector, as well as make comparisons to the other industries.

### Data Preparation

As mentioned above, we have three columns of data that are essential for the model: Employed, Retail, and COVID. They were not variables available from the raw data so we had to do a few mutations.

* Employed - We are able to identify an individual’s employment status using the “EMPSTAT” column from the raw data. However, this variable is categorical, with various codes representing different employment statuses such as “Has a job, not at work last week” or “Unemployed, new worker”.  To simplify our column’s variables we categorized the different employment status codes so our column directly answers the question; Is the person currently working or not? We also chose to filter out any individuals who were not part of the labor force.  That way, for individuals who are marked as unemployed, we can assume they are actively seeking work.
* Retail - By using the “IND” column from the raw data, we were able to identify individuals that work in the retail industry. Based on this information, we created the “retail” column that assigns a value of “true” or “false” depending on whether that individual worked in the retail industry.
* COVID - We used the date column to create the “COVID” column that indicates whether an observation was taken during or after the COVID-19 pandemic.

data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

#prepare data  
data <- data %>% filter( YEAR >= 2018 ) %>%   
 filter(!(EMPSTAT == 1 | EMPSTAT >= 30 )) %>%  
 mutate(employed = ifelse(EMPSTAT == 10 | EMPSTAT == 12, TRUE, FALSE))  
  
df <- inner\_join(data, ind, by = c('IND' = 'ind'))  
  
df <- df %>% mutate(retail = ifelse(indname == 'Retail Trade', TRUE, FALSE)) %>%  
 mutate(date = as.Date(paste(YEAR, MONTH, "01", sep = "-"))) %>%   
 mutate(COVID = ifelse(date >= as.Date("2020-03-01"), TRUE, FALSE))

### A Glimpse of the Data

Before getting our model set up, we wanted get a general idea of how COVID-19 effected the labor force. To do this, we graphed the total percentage of the labor force each industry contributed overtime. Surprisingly, our graph shows the retail industry does not seem to be effected by COVID-19 as much as we expected. This is especially true when its compared to the entertainment & hospitality industry. Nonetheless, we will dig deeper on this subject with our model.

graphdf <- df %>% group\_by(indname, date) %>% summarize(totalworkers = sum(employed == TRUE))

`summarise()` has grouped output by 'indname'. You can override using the  
`.groups` argument.

bymonthlbr <- graphdf %>% group\_by(date) %>% summarize(lbrforce = sum(totalworkers))  
  
graphdf <- merge(graphdf, bymonthlbr, by = "date")  
  
graphdf <- graphdf %>% mutate(labrforcepercent = round((totalworkers/lbrforce) \* 100, 2))  
  
ggplot(graphdf, aes(x = date, y = labrforcepercent, color = indname)) +  
 geom\_line(size = 1) +  
 ggtitle("Percentage of Labor Force by Industry") +  
 labs(x = "Date", y = "Labor Force Percentage", color = "Industry") +  
 geom\_vline(xintercept = as.numeric(as.Date("2020-03-01")), color = "red", linetype = "dashed") +  
 geom\_label(aes(x = as.Date("2020-03-01"), y = 17, label = "COVID-19 Lockdowns Begin"),  
 fill = "white", color = "red", size = 4, fontface = "bold", vjust = -0.5, hjust = -.1) +   
 theme(panel.grid.major.x = element\_blank(),  
 panel.grid.major.y = element\_line(color = "gray", size = 0.5),  
 panel.grid.minor = element\_blank())

Warning: The `size` argument of `element\_line()` is deprecated as of ggplot2 3.4.0.  
ℹ Please use the `linewidth` argument instead.



### Model Controls

For our model, we identified the following controls:

* Sex - An individual’s sex may play a role in whether one is hired or not. Even though it is illegal to discriminate based on sex or gender.
* Age - Another factor that could lead to being discriminated against. (Also illegal)
* Race - Similar to sex and age, an individual could be discriminated against for their race. (Also illegal)
* Seasonality - Retail companies would often hire seasonal workers to have smoother operations during their peak seasons.

#set up model, look at initial coefficients  
model <-feglm(employed ~ retail \* COVID + i(MONTH) + AGE + SEX + RACE, family = binomial(link = 'logit'), data = df, se = 'hetero')  
  
etable(model, drop = 'MONTH')

model  
Dependent Var.: employed  
   
Constant 2.534\*\*\* (0.0176)  
retailTRUE -0.2960\*\*\* (0.0132)  
COVIDTRUE -0.5333\*\*\* (0.0064)  
AGE 0.0210\*\*\* (0.0003)  
SEX 0.0443\*\*\* (0.0058)  
RACE -0.0004\*\*\* (1.52e-5)  
retailTRUE x COVIDTRUE 0.1390\*\*\* (0.0174)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedasti.-rob.  
Observations 2,847,469  
Squared Cor. 0.00733  
Pseudo R2 0.01789  
BIC 1,010,120.7  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Additional Calculations for Binary Independent Variables & Interactions

Logit interaction models are special in that the interaction coefficient of a model does not capture the magnitude of the effect we are actually looking for. In order to properly find the effect of the interaction, we will have to find the cross derivative. We achieve this by calculating the average marginal effect when COVID is False, and subtract it from the average marginal effect when COVID is True. For non-interaction terms, instead of the coefficient found by the model we use the average marginal effects.

#find average marginal effects  
avg\_slopes(model, variables = c('retail','COVID'))

Term Contrast Estimate Std. Error z Pr(>|z|) S 2.5 % 97.5 %  
 COVID TRUE - FALSE -0.02140 0.000249 -86.1 <0.001 Inf -0.0219 -0.02092  
 retail TRUE - FALSE -0.00971 0.000420 -23.1 <0.001 389.7 -0.0105 -0.00888  
  
Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high   
Type: response

#calculate cross derivative  
slopes(model, datagrid(COVID = FALSE, grid.type = 'counterfactual'), variables = c('retail','COVID'))

Warning: Some of the variable names are missing from the model data: grid.type

Term Contrast COVID grid.type Estimate Std. Error z Pr(>|z|)  
 COVID TRUE - FALSE FALSE counterfactual -0.0224 0.000343 -65.1 <0.001  
 retail TRUE - FALSE FALSE counterfactual -0.0111 0.000554 -20.0 <0.001  
 S 2.5 % 97.5 %  
 Inf -0.0230 -0.02170  
 293.0 -0.0122 -0.00998  
  
Columns: rowid, term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, COVID, grid.type, predicted\_lo, predicted\_hi, predicted, employed, retail, MONTH, AGE, SEX, RACE   
Type: response

slopes(model, datagrid(COVID = TRUE, grid.type = 'counterfactual'),variables = c('retail','COVID'))

Warning: Some of the variable names are missing from the model data: grid.type

Term Contrast COVID grid.type Estimate Std. Error z Pr(>|z|)  
 COVID TRUE - FALSE TRUE counterfactual -0.0224 0.000343 -65.1 <0.001  
 retail TRUE - FALSE TRUE counterfactual -0.0089 0.000689 -12.9 <0.001  
 S 2.5 % 97.5 %  
 Inf -0.0230 -0.02170  
 124.6 -0.0103 -0.00755  
  
Columns: rowid, term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, COVID, grid.type, predicted\_lo, predicted\_hi, predicted, employed, retail, MONTH, AGE, SEX, RACE   
Type: response

interaction <- (-.0111 + (-.0089))  
interaction

[1] -0.02

### Results + Interpretation

We are now able to interpret our model:

* The COVID-19 pandemic, during or after, is associated with a 0.02140 decrease in the probability of being employed
* Working in the retail industry is associated with a 0.00971 decrease in the probability of being e
* The effect of the COVID-19 pandemic is associated with a .0022 decrease in the probability that a person in the retail industry is employed.

With consideration that COVID has negatively impacted the probability of employment across all industries substantially, the retail industry appears to have been relatively resilient. Although the retail industry already exhibited a negative effect on employment probability before COVID, the additional decrease of .0022 in probability seems minimal in comparison.

This could be attributed to the wide range of business types that fall within the industry. For instance, grocery stores, which are classified as retail, remained essential during the pandemic due to the perpetual demand for their products. Additionally, non-essential retail businesses also found demand during the pandemic. For example, many individuals saw lock-down as an opportunity to make home improvements, and found themselves purchasing goods from home-improvement stores such as IKEA. The pandemic was also encouraged individuals to explore new hobbies, which could result in more spending.

It’s crucial to also consider that COVID may have simply redirected shopping habits from in-person to online channels. Nevertheless, retail businesses managed to generate profits, adapting to the changing landscape through various means.

## Research Question 3: Retail needs to worry about who has money to spend - what has changed about who is working and earning money?

Given the research question “Retail needs to worry about who has money to spend - what has changed about who is working and earning money?”This analysis aims to dissect the dynamics of employment and income in the context of potentially shifting consumer and purchase power bases for the retail industry, especially in light of the COVID-19 pandemic. To answer the questions thoroughly, the analysis is structured to examine both individual (people level) and sector-wide (industry level) perspectives. This approach will enable a detailed exploration of how employment status, income variations, and industry-specific trends collectively impact consumer spending and the income market landscape.

To understand how COVID-19 has affected people’s earnings, a Different in Different model is employed to reveal whether individuals experienced a change in income before and after COVID. This model uses independent variables such as demographic variables (age, sex and race) along with employment status to measure the impact on income across different groups. The “Retail Trade” industry is designated as the treatment group, which allows for a direct assessment of how this sector has been impacted compared to others. We utilize the post-treatment variable to distinguish between the periods before and after COVID-19. This allows for comparisons between pre-COVID and pandemic era periods.

Visualization

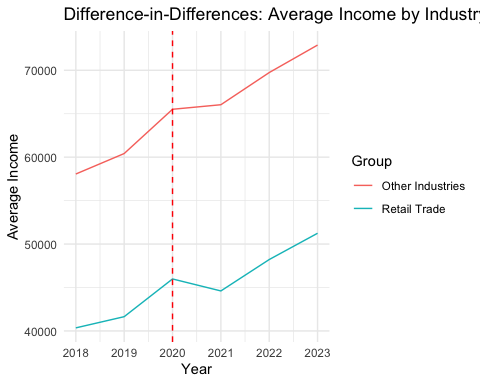
df\_final$EMPSTAT <- ifelse(df\_final$EMPSTAT >= 20, "Unemployed", "Employed")  
unique(df\_final$RACE)

<labelled<integer>[26]>: Race  
 [1] 100 801 651 802 652 200 805 806 300 803 804 810 809 814 816 807 830 813 818  
[20] 811 808 812 820 819 817 815  
  
Labels:  
 value label  
 100 White  
 200 Black  
 300 American Indian/Aleut/Eskimo  
 650 Asian or Pacific Islander  
 651 Asian only  
 652 Hawaiian/Pacific Islander only  
 700 Other (single) race, n.e.c.  
 801 White-Black  
 802 White-American Indian  
 803 White-Asian  
 804 White-Hawaiian/Pacific Islander  
 805 Black-American Indian  
 806 Black-Asian  
 807 Black-Hawaiian/Pacific Islander  
 808 American Indian-Asian  
 809 Asian-Hawaiian/Pacific Islander  
 810 White-Black-American Indian  
 811 White-Black-Asian  
 812 White-American Indian-Asian  
 813 White-Asian-Hawaiian/Pacific Islander  
 814 White-Black-American Indian-Asian  
 815 American Indian-Hawaiian/Pacific Islander  
 816 White-Black--Hawaiian/Pacific Islander  
 817 White-American Indian-Hawaiian/Pacific Islander  
 818 Black-American Indian-Asian  
 819 White-American Indian-Asian-Hawaiian/Pacific Islander  
 820 Two or three races, unspecified  
 830 Four or five races, unspecified  
 999 Blank

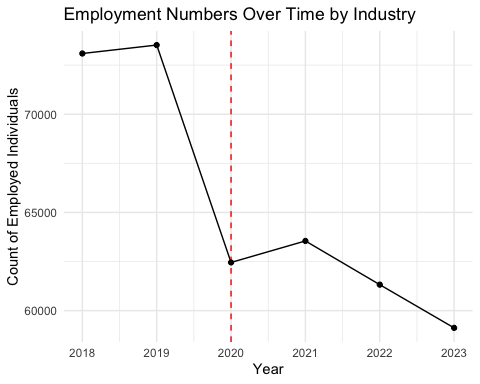
#Create a plot to compare the average income between Retail Trade and Other Industries  
df\_final$group <- ifelse(df\_final$indname == "Retail Trade", "Retail Trade", "Other Industries")  
  
  
avg\_income\_by\_group\_year <- df\_final %>%  
 group\_by(group, YEAR) %>%  
 summarise(mean\_income = mean(INCTOT, na.rm = TRUE)) %>%  
 ungroup()

`summarise()` has grouped output by 'group'. You can override using the  
`.groups` argument.

# Avergae Income Plot  
ggplot(avg\_income\_by\_group\_year, aes(x = YEAR, y = mean\_income, color = group)) +  
 geom\_line() +  
 geom\_vline(xintercept = 2020, linetype="dashed", color = "red")+  
 labs(title = "Difference-in-Differences: Average Income by Industry",  
 x = "Year",  
 y = "Average Income",  
 color = "Group") +  
 theme\_minimal()



# Calculate employment counts by group and year  
employment\_counts\_by\_group\_year <- df\_final %>%  
 filter(EMPSTAT == "Employed") %>%  
 group\_by(YEAR) %>%  
 summarise(count\_employed = n(), .groups = 'drop') %>%  
 ungroup()  
  
# Plot  
ggplot(employment\_counts\_by\_group\_year, aes(x = YEAR, y = count\_employed)) +  
 geom\_line() +  
 geom\_vline(xintercept = 2020, linetype="dashed", color = "red")+  
 geom\_point() + # Adding points for clarity  
 labs(title = "Employment Numbers Over Time by Industry",  
 x = "Year",  
 y = "Count of Employed Individuals",  
 color = "Group") +  
 theme\_minimal()



This graph  likely shows a downward trend in the number of employed individuals over time. Both Pre Covid and post Covid have affected the employment market negatively. Consequently, numerous individuals have lost their jobs or faced layoffs. This leads to a potential decrease in spending power. Although there is a noted increase in average income after the onset of COVID, this growth must be contextualized with the rising inflation rates, economic challenges, and shortages of food and goods, all of which pose significant obstacles to consumer spending capabilities.

df\_final <- df\_final %>%  
 mutate(post\_treatment = YEAR >= 2020) %>%  
 group\_by(CPSIDV, YEAR) %>%  
 mutate(mean\_income = mean(INCTOT, na.rm = TRUE)) %>% ungroup()  
  
df\_final$SEX <- ifelse(df\_final$SEX == 1, "Male", "Female")  
df\_final$RACE <- ifelse(df\_final$RACE == 100, "White",  
 ifelse(df\_final$RACE == 200, "Black",  
 ifelse(df\_final$RACE == 651, "Asian", "Other")))  
  
  
  
#Diff in Diff Model  
  
df\_final$treated\_group <- ifelse(df\_final$indname == "Retail Trade", 1, 0)  
  
# Run the DiD model with an interaction between treated\_group and post\_treatment  
dif\_in\_dif\_model <- feols(INCTOT ~ treated\_group \* post\_treatment + AGE + SEX + factor(RACE) + factor(EMPSTAT)| SERIAL, data = df\_final)  
  
# View the summary of the model  
etable(dif\_in\_dif\_model)

dif\_in\_dif\_model  
Dependent Var.: INCTOT  
   
treated\_group -10,564.2\*\*\* (602.0)  
post\_treatmentTRUE 9,280.1\*\*\* (329.8)  
AGE 1,271.1\*\*\* (11.72)  
SEXMale 21,379.8\*\*\* (243.8)  
factor(RACE)Black -25,138.1\*\*\* (878.0)  
factor(RACE)Other -21,336.9\*\*\* (988.5)  
factor(RACE)White -11,865.3\*\*\* (782.1)  
factor(EMPSTAT)Unemployed -31,255.3\*\*\* (452.9)  
treated\_group x post\_treatmentTRUE -3,181.0\*\*\* (785.1)  
Fixed-Effects: --------------------  
SERIAL Yes  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E.: Clustered by: SERIAL  
Observations 413,016  
R2 0.31391  
Within R2 0.07391  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

According to the table, there are statistically significant relationships between various factors and income changes, before and after the pandemic. Employees in the retail sector experienced an average income reduction of $10,564 compared to the other industries, before the pandemic. This indicates that the retail employee has the pre-pandemic challenges faced. After the COVID, the income increased by $9,280 for all industries, suggesting an overall positive effect of COVID impact on income. Furthermore, demographics act as a positive determinant of income. For each additional year of age, the income increases by $1,271, holding all else constant. Male individuals earn significantly more, with an average income increase of $21,379.8 compared to females. In terms of racial comparisons with the Asian demographic, Black individuals earn $25,138.1 less, those classified as Other $21,336.9 less, and White individuals $11,865.3 less. Yet, being unemployed compared to the employed baseline category is associated with a $31,255 decrease in income, highlighting the severe economic impact of job loss. On average, individuals in the retail sector experienced an additional income decrease of $3,181 after COVID-19 compared to those in other sectors.

# **Changes to money holders by industry:**

In the wake of the COVID-19 pandemic, there have been significant changes to the makeup of the US labor force. Looking at the rate of labor force participation across all industries shows statistically significant drops not just in retail but across industries.

data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

#factor years,   
data <- data %>% mutate(data, yf = factor(data$YEAR))  
  
#precovid and post covid for diff in diff approaches  
data <- data %>% mutate(data, pndmc = case\_when( YEAR < 2020 | (YEAR == 2020 & MONTH < 4) ~ 0,  
 TRUE ~ 1))  
#factor sex  
data <- data %>% mutate(data, sx = case\_when(SEX < 2 ~ 0,  
 TRUE ~ 1))  
  
  
  
  
#Industry remapping to make workable, took 280 different indusrty codes and condensed into 16, also truncated names for nicer graphing  
data <- data %>% mutate(data, indclust = case\_when( IND < 1 ~ "NIU",  
 IND < 300 ~ "AG & Frstry",  
 IND < 1000 ~ "Mining, Util, & Const",  
 IND < 1400 ~ "Secondary Food",  
 IND < 4000 ~ "Manufacturing",  
 IND < 4600 ~ "Wholesale",  
 IND < 4700 | (IND > 8769 & IND < 8781) ~ "Auto",  
 IND < 4900 ~ "Furn + Appliances",  
 IND < 5000 ~ "Food Sellers",  
 IND < 5085 | IND == 7480 | (IND > 7900 & IND < 8300) ~ "Healthcare",  
 IND < 6000 ~ "Retail",  
 IND < 6400 ~ "Transport",  
 IND < 6800 | (IND > 8560 & IND < 8600) ~ "Info & entertainment",  
 IND < 7075 | (IND > 8659 & IND < 8671) ~ "Finance",  
 (IND > 7800 & IND < 7900) ~ "Education",  
 TRUE ~ "Other"  
   
 ))  
  
  
  
  
  
#bianary of unable to work due to covid  
data <- data %>% mutate(data, cvunemp = case\_when(COVIDUNAW == 1 ~ 0,   
 COVIDUNAW == 2 ~ 1  
 ) )  
  
  
#factor and filter in labor force data  
labdata <- data %>% filter(data$LABFORCE > 0)  
  
  
labdata <- labdata %>% mutate(labdata, labf2 = case\_when(LABFORCE == 1 ~ 0,  
 TRUE ~ 1))  
  
  
  
  
  
  
incdata <- data %>% filter(!is.na(INCTOT))

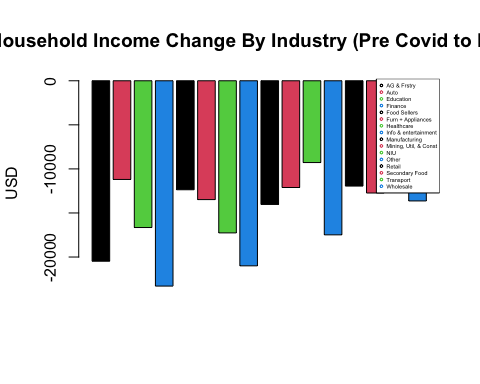
One assumption made in this analysis is that the industry a person works in is on average likely to be the field they continue to work in. As such the aggregate changes shown at the industry level can give practical insights to retailers about the changing income distributions of Americans. Another assumption is that industry income data is useful to retailers as industries likely have a high correlation with the kind of highly tailored advertising used across online retail. When controlling for gender there were statically significant drops in laborforce participation in transportation, wholesalers, entertainment, and furniture/appliance sellers. Each of these industries saw a drop of about 0.3% compared to pre pandemic participation levels.

labforcepr\_feols <- feols(labf2 ~ pndmc + indclust + (indclust\*pndmc), data = labdata)  
  
labforceprsx\_feols <- feols(labf2 ~ pndmc + indclust + (indclust\*pndmc) + sx, data = labdata)  
  
  
etable(labforcepr\_feols, labforceprsx\_feols, vcov = 'hetero')

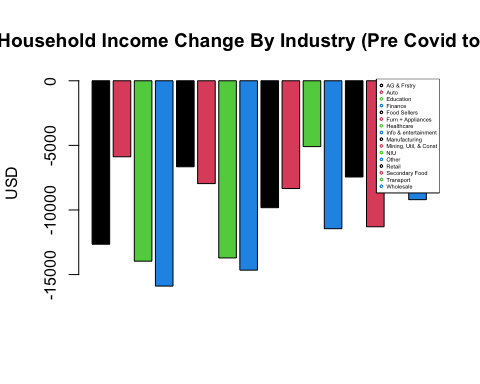
labforcepr\_feols labforceprsx\_feols  
Dependent Var.: labf2 labf2  
   
Constant 0.9874\*\*\* (0.0007) 0.9885\*\*\* (0.0007)  
pndmc 0.0003 (0.0010) 0.0003 (0.0010)  
indclustAuto 0.0060\*\*\* (0.0008) 0.0056\*\*\* (0.0008)  
indclustEducation 0.0041\*\*\* (0.0007) 0.0059\*\*\* (0.0007)  
indclustFinance 0.0063\*\*\* (0.0007) 0.0075\*\*\* (0.0007)  
indclustFoodSellers -7.11e-5 (0.0009) 0.0009 (0.0009)  
indclustFurn+Appliances 0.0042\*\*\* (0.0009) 0.0045\*\*\* (0.0009)  
indclustHealthcare 0.0064\*\*\* (0.0007) 0.0085\*\*\* (0.0007)  
indclustInfo&entertainment 0.0005 (0.0008) 0.0012 (0.0008)  
indclustManufacturing 0.0074\*\*\* (0.0007) 0.0074\*\*\* (0.0007)  
indclustMining,Util,&Const 0.0066\*\*\* (0.0007) 0.0060\*\*\* (0.0007)  
indclustNIU -0.9778\*\*\* (0.0007) -0.9762\*\*\* (0.0007)  
indclustOther 0.0032\*\*\* (0.0007) 0.0042\*\*\* (0.0007)  
indclustRetail -0.0014. (0.0008) -8.43e-5 (0.0008)  
indclustSecondaryFood 0.0043\*\*\* (0.0009) 0.0048\*\*\* (0.0009)  
indclustTransport 0.0044\*\*\* (0.0008) 0.0044\*\*\* (0.0008)  
indclustWholesale 0.0086\*\*\* (0.0008) 0.0087\*\*\* (0.0008)  
pndmc x indclustAuto -0.0005 (0.0012) -0.0005 (0.0012)  
pndmc x indclustEducation -0.0003 (0.0011) -0.0003 (0.0011)  
pndmc x indclustFinance 1.69e-5 (0.0011) -6.24e-5 (0.0011)  
pndmc x indclustFoodSellers -0.0022 (0.0014) -0.0022 (0.0014)  
pndmc x indclustFurn+Appliances -0.0026. (0.0014) -0.0027\* (0.0014)  
pndmc x indclustHealthcare -0.0017 (0.0011) -0.0018 (0.0011)  
pndmc x indclustInfo&entertainment -0.0034\*\* (0.0013) -0.0034\*\* (0.0013)  
pndmc x indclustManufacturing -0.0014 (0.0011) -0.0014 (0.0011)  
pndmc x indclustMining,Util,&Const -0.0017 (0.0011) -0.0017 (0.0011)  
pndmc x indclustNIU -0.0017 (0.0011) -0.0018. (0.0011)  
pndmc x indclustOther -0.0021. (0.0011) -0.0021. (0.0011)  
pndmc x indclustRetail -0.0016 (0.0012) -0.0017 (0.0012)  
pndmc x indclustSecondaryFood -0.0007 (0.0014) -0.0008 (0.0014)  
pndmc x indclustTransport -0.0024\* (0.0012) -0.0024\* (0.0012)  
pndmc x indclustWholesale -0.0030\* (0.0012) -0.0030\* (0.0012)  
sx -0.0041\*\*\* (0.0001)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob. Heteroskedast.-rob.  
Observations 3,657,067 3,657,067  
R2 0.94902 0.94904  
Adj. R2 0.94902 0.94904  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

This industry variation also shows up strongly in the decreases in both mean and median household income but with some differences from the changes shown in labor force participation. While all industries saw decreases in both mean and median household income after the pandemic, agriculture, finance, entertainment, and surprisingly healthcare have been among the industries hardest hit.

#Industry mean HHIncome  
  
mean\_hhincome <- incdata %>% group\_by(indclust) %>% summarize(mean\_HHIncome = mean(HHINCOME, na.rm = TRUE))  
  
precinc <- incdata %>% filter(incdata$pndmc < 1)  
  
mean\_hhincomeprec <- precinc %>% group\_by(indclust) %>% summarize(mean\_HHIncome = mean(HHINCOME, na.rm = TRUE))  
  
postcinc <- incdata %>% filter(incdata$pndmc == 1)  
  
mean\_hhincomepostc <- postcinc %>% group\_by(indclust) %>% summarize(mean\_HHIncome = mean(HHINCOME, na.rm = TRUE))  
  
  
barplot(mean\_hhincomeprec$mean\_HHIncome - mean\_hhincomepostc$mean\_HHIncome, col= 1:4, main = "Mean Household Income Change By Industry (Pre Covid to Post Covid)", ylab = "USD")  
legend('topright', legend= mean\_hhincome$indclust,  
 col=1:4, pch=1, cex=0.35)

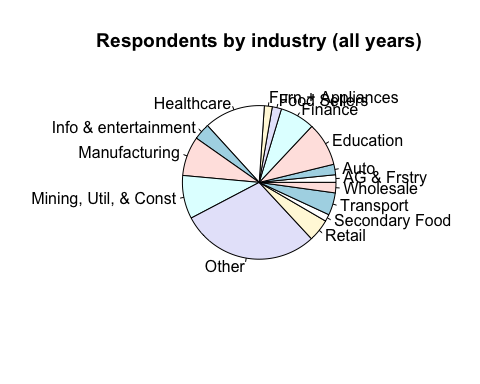


median\_hhincomeprec <- precinc %>% group\_by(indclust) %>% summarize(median\_HHIncome = median(HHINCOME, na.rm = TRUE))  
  
median\_hhincomepostc <- postcinc %>% group\_by(indclust) %>% summarize(median\_HHIncome = median(HHINCOME, na.rm = TRUE))  
  
barplot(median\_hhincomeprec$median\_HHIncome - median\_hhincomepostc$median\_HHIncome, col= 1:4, main = "Median Household Income Change By Industry (Pre Covid to Post Covid)", ylab = "USD")  
legend('topright', legend= mean\_hhincome$indclust,  
 col=1:4, pch=1, cex=0.35)



Among respondents to the census covid survey, there were significantly different levels of persons reporting being unable to work due to covid. Retail had a relatively high rate of around 7% of respondents reporting being unable to work because of Covid. The info and entertainment industry again stands out as the hardest hit with 11.92% of respondents reporting being unable to work. However, the regression alone does not tell us about the present employment of these respondents and whether they have been able to return to work. Additionally, combining this data with the other industry-based regression paints a picture of the covid pandemic having a depreciating effect on labor force participation with variation across industries. The impacts of the pandemic have been felt across all industries and demographics but the impacts have not been equal for all.

lniu <- labdata %>% filter(indclust != 'NIU')  
  
pie(table(lniu$indclust), main = "Respondents by industry (all years)")



ldf <- lniu %>% filter(pndmc < 1)  
  
  
  
fcvid <- feols(cvunemp ~ indclust, data = labdata)

NOTE: 2,328,119 observations removed because of NA values (LHS: 2,328,119).

fcvid2 <- feols(cvunemp ~ indclust + LABFORCE, data = labdata)

NOTE: 2,328,119 observations removed because of NA values (LHS: 2,328,119).

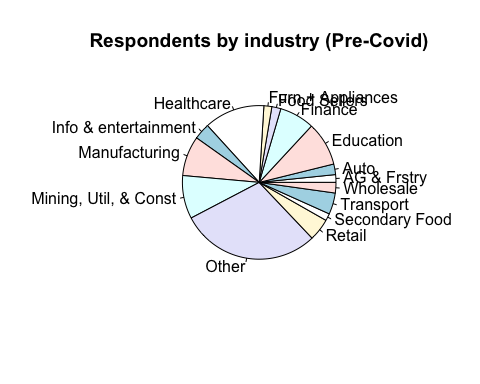
fcvid3 <- feols(cvunemp ~ indclust + factor(LABFORCE) + (indclust\*factor(LABFORCE)), data = labdata)

NOTE: 2,328,119 observations removed because of NA values (LHS: 2,328,119).

etable(fcvid, fcvid2)

fcvid fcvid2  
Dependent Var.: cvunemp cvunemp  
   
Constant 0.0379\*\*\* (0.0018) 0.1698\*\*\* (0.0045)  
indclustAuto 0.0187\*\*\* (0.0023) 0.0191\*\*\* (0.0023)  
indclustEducation 0.0118\*\*\* (0.0019) 0.0120\*\*\* (0.0019)  
indclustFinance 0.0132\*\*\* (0.0020) 0.0136\*\*\* (0.0020)  
indclustFoodSellers -0.0023 (0.0024) -0.0025 (0.0024)  
indclustFurn+Appliances 0.0097\*\*\* (0.0025) 0.0098\*\*\* (0.0025)  
indclustHealthcare 0.0042\* (0.0019) 0.0045\* (0.0019)  
indclustInfo&entertainment 0.0813\*\*\* (0.0021) 0.0810\*\*\* (0.0021)  
indclustManufacturing 0.0090\*\*\* (0.0019) 0.0093\*\*\* (0.0019)  
indclustMining,Util,&Const 0.0318\*\*\* (0.0019) 0.0321\*\*\* (0.0019)  
indclustNIU -0.0048\*\* (0.0018) -0.0698\*\*\* (0.0027)  
indclustOther 0.0372\*\*\* (0.0018) 0.0372\*\*\* (0.0018)  
indclustRetail 0.0298\*\*\* (0.0021) 0.0296\*\*\* (0.0021)  
indclustSecondaryFood 0.0086\*\*\* (0.0026) 0.0088\*\*\* (0.0026)  
indclustTransport 0.0291\*\*\* (0.0020) 0.0292\*\*\* (0.0020)  
indclustWholesale 0.0082\*\*\* (0.0023) 0.0086\*\*\* (0.0023)  
LABFORCE -0.0663\*\*\* (0.0021)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type IID IID  
Observations 1,328,948 1,328,948  
R2 0.00718 0.00795  
Adj. R2 0.00716 0.00794  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

pie(table(ldf$indclust), main = "Respondents by industry (Pre-Covid)")



lpdf <- lniu %>% filter(pndmc > 0)  
  
pie(table(lpdf$indclust), main = "Respondents by industry (Post-Covid)")

