Sta303- Professional Development

```
library("tidytuesdayR")
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
tuesdata <- tidytuesdayR::tt_load('2021-02-23')
```

```
## --- Compiling #TidyTuesday Information for 2021-02-23 ----
```

```
## --- There are 2 files available ---
```

```
## --- Starting Download ---
```

```
##
##  Downloading file 1 of 2: `earn.csv`
##  Downloading file 2 of 2: `employed.csv`
```

```
## --- Download complete ---
```

```
earn<- tuesdata$earn
employed <- tuesdata$employed

employed %>%
  count(race_gender, sort = TRUE)
```

```
## # A tibble: 6 x 2
##   race_gender                 n
##   <chr>                   <int>
```

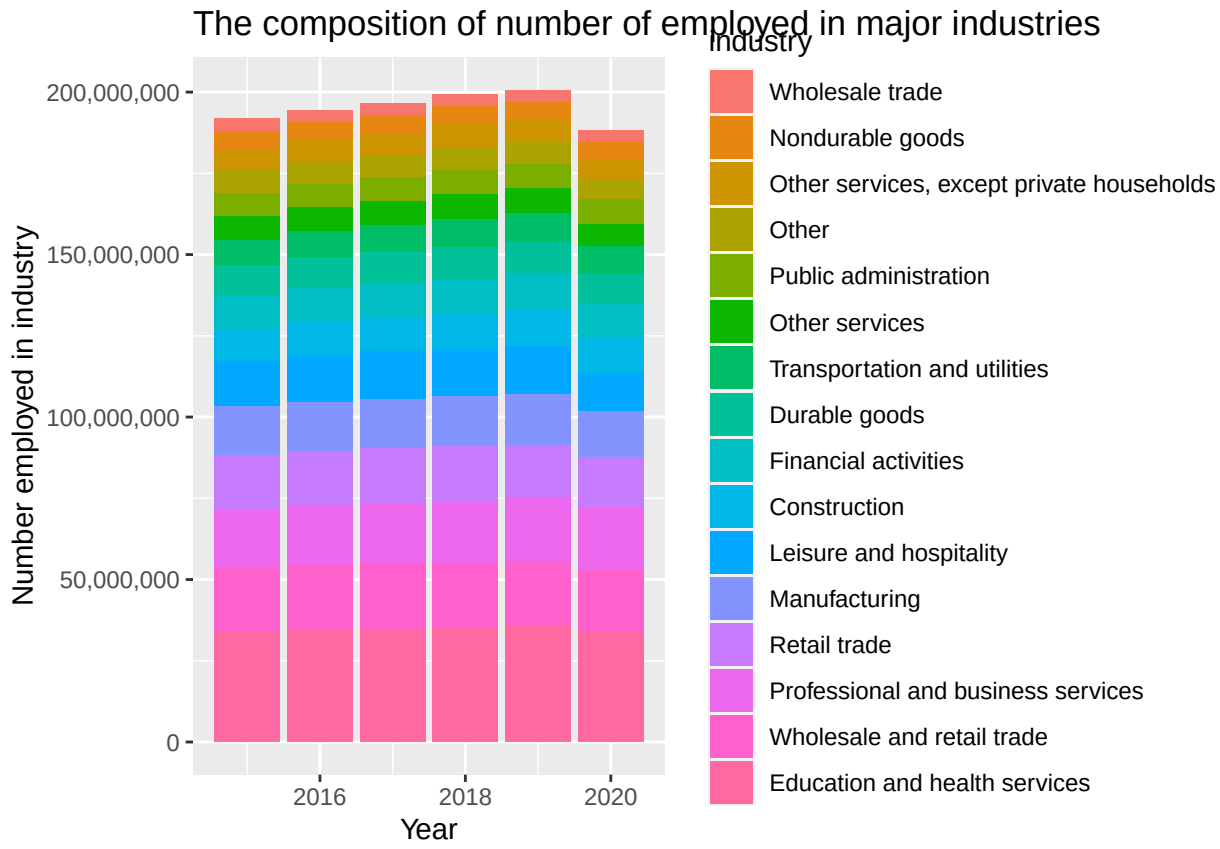
```
## 1 Black or African American  1386
## 2 Men                        1386
## 3 TOTAL                      1386
## 4 White                      1386
## 5 Women                      1386
## 6 Asian                      1254
```

```r
# The original data does not categorize gender and race
# We need to specify the categories

employed<- tuesdata$employed %>%
  mutate(dimension = case_when(
    race_gender == "TOTAL" ~ "Total",
    race_gender %in% c("Men", "Women") ~ "Gender",
    TRUE ~ "Race"
  ))
```

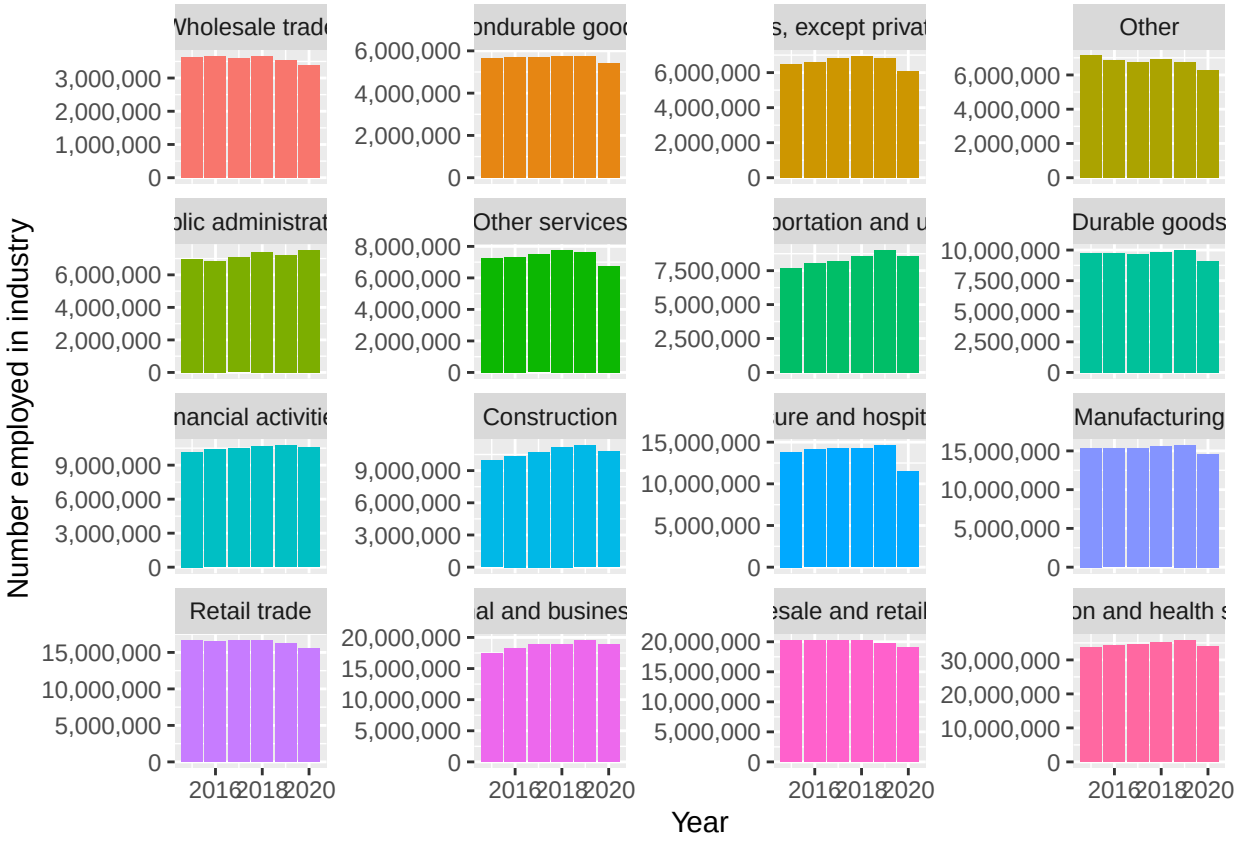
Bar plot that shows the composition of number of employed in major industries.

```
employed_clean<- employed %>%
  filter(!is.na(employ_n)) %>%
  mutate(industry = fct_lump(industry, 15, w = employ_n), # The top 15 industries with largest workers.
         industry = fct_reorder(industry, employ_n, sum)) # Aggregate the workers in each industries.
# Bar plot
employed_clean %>%
  filter(dimension == "Total") %>%
  ggplot(aes(x = year, y = employ_n, fill = industry)) +
  geom_col() +
  scale_y_continuous(labels = comma) +
  labs(y = "Number employed in industry",
       x = "Year",
       title = "The composition of number of employed in major industries")
```



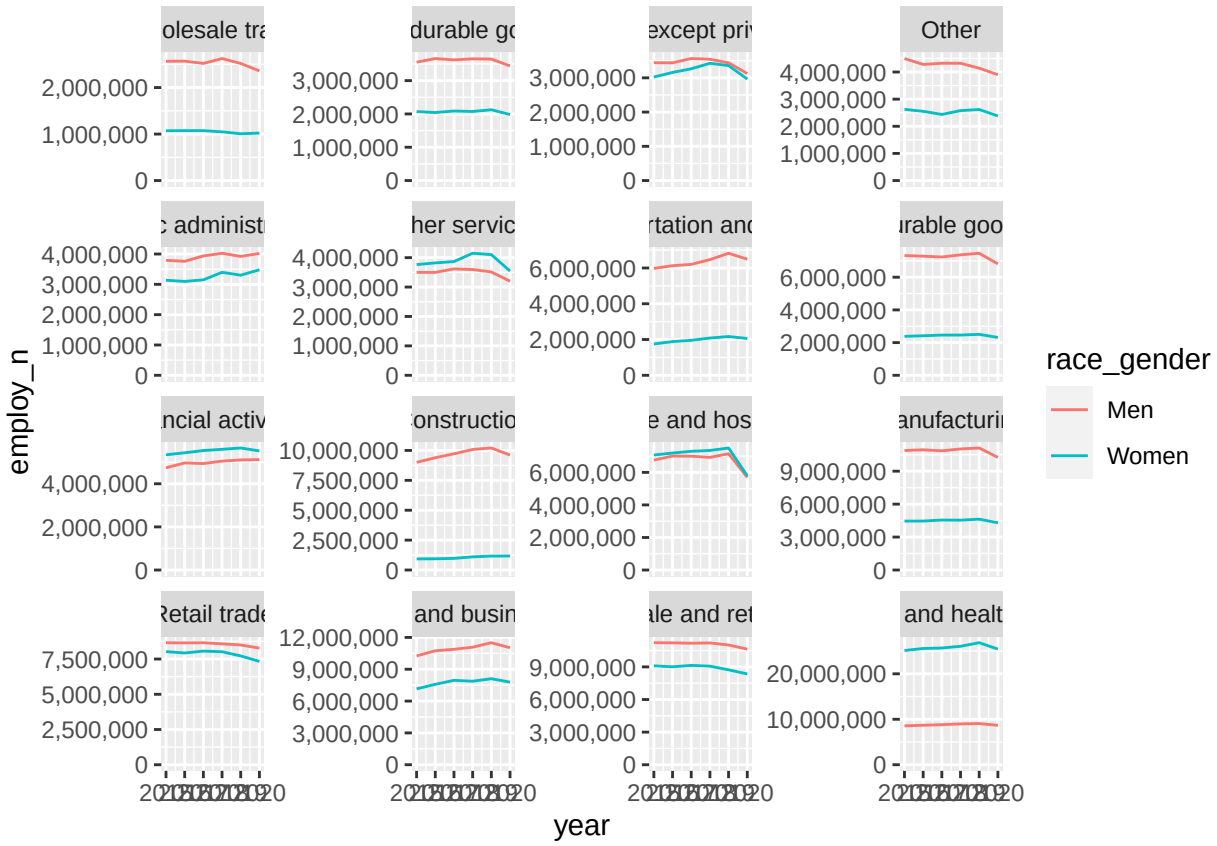
Facet bar plots that show the total number of workers in each industry for each year.

```
# Facet bar plots that show the total number of workers in each industry for each year.
employed_clean %>%
  filter(dimension == "Total") %>%
  ggplot(aes(x = year, y = employ_n, fill = industry)) +
  geom_col() +
  facet_wrap(~industry, scales = "free_y") +
  theme(legend.position =  "none") +
  scale_y_continuous(labels = comma) +
  labs(y = "Number employed in industry",
       x = "Year")
```



```
employed_gender <- employed_clean %>%
  filter(dimension == "Gender") %>%
  filter(!is.na(employ_n)) %>%
  group_by(industry, year, race_gender)%>%
  summarise(employ_n = sum(employ_n), .groups = "drop")

employed_gender %>%
  ggplot(aes(x = year, y = employ_n, color = race_gender)) +
  geom_line() +
  facet_wrap(~industry, scales = "free_y") +
  scale_y_continuous(labels = comma) +
  expand_limits(y = 0)
```




```
  labs(y = "Number employed in industry",
       x = "Year")
```

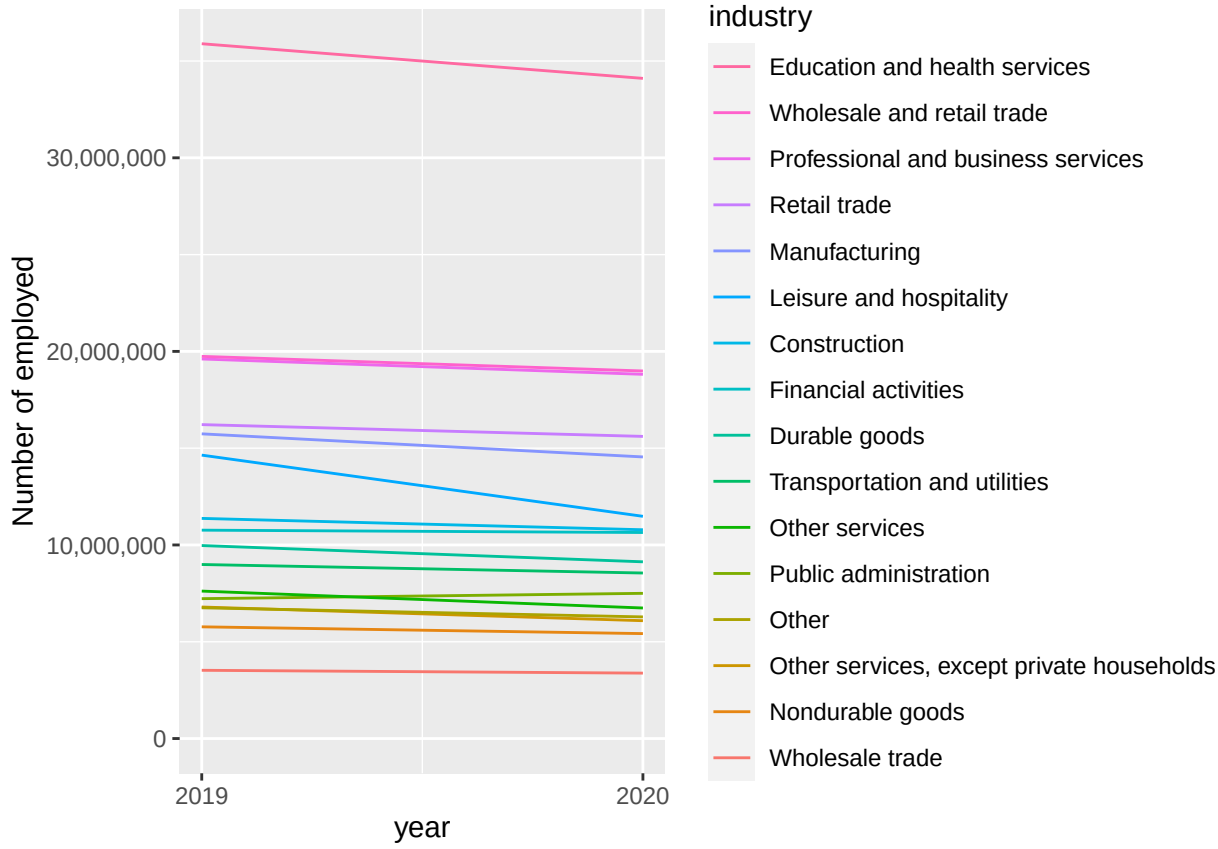
```
## $y
## [1] "Number employed in industry"
##
## $x
## [1] "Year"
##
## attr(,"class")
## [1] "labels"
```

Since there is not so much variation in employment between male and female from 2015 to 2019. Let's look at 2019 to 2020 specifically.

```
industry_2019_2020 <- employed_clean %>%
  filter(year %in% c(2019, 2020)) %>%
  group_by(year, industry, dimension, race_gender) %>%
  summarise(employ_n = sum(employ_n), .groups = "drop")
```

A line plot that shows the number of employment from 2019 to 2020 in each industry.

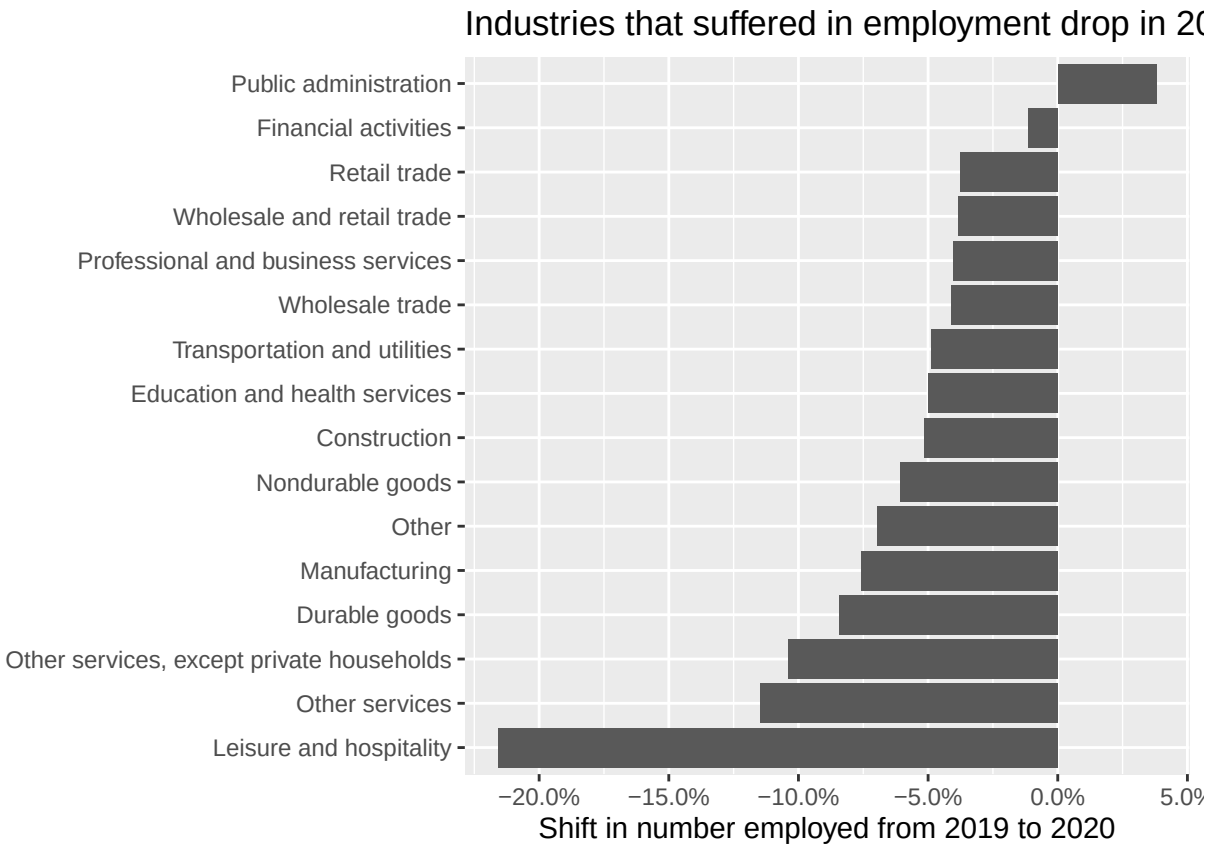
```
industry_2019_2020 %>%
  filter(dimension == "Total") %>%
  ggplot(aes(x = year, y = employ_n, color = industry)) +
  geom_line() +
  expand_limits(y = 0) +
  scale_y_continuous(labels = comma) +
  scale_x_continuous(breaks = 2019:2020) +
  scale_color_discrete(guide = guide_legend(reverse = TRUE)) +
  labs(x = "year",
       y = "Number of employed")
```



Estimate the effect of pandemic on employment for each industry

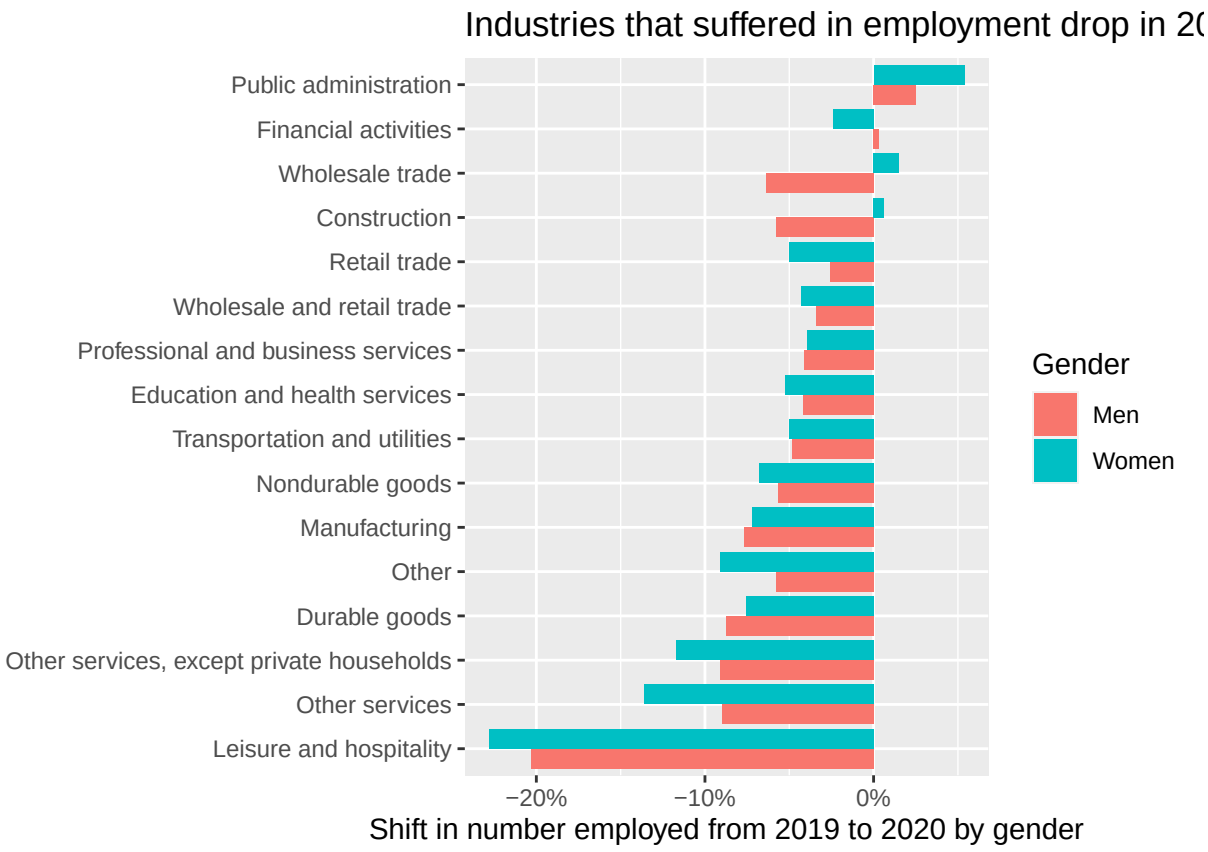
```
compare_2019_2020 <- industry_2019_2020 %>%
  arrange(year) %>%
  group_by(industry, dimension, race_gender) %>%
  summarize(ratio = last(employ_n)/ first(employ_n),
            change = ratio - 1,
            employed_2019 = first(employ_n), .groups = "drop") %>%
  ungroup()

compare_2019_2020 %>%
  filter(dimension == "Total") %>%
  mutate(industry = fct_reorder(industry,change)) %>%
  ggplot(aes(x = change, y = industry)) +
  geom_col() +
  scale_x_continuous(labels = percent) +
  labs(title = "Industries that suffered in employment drop in 2020",
       x = "Shift in number employed from 2019 to 2020",
       y = "")
```



Estimate the effect of pandemic on employment for each industry on gender.

```
compare_2019_2020 %>%
  filter(dimension == "Gender") %>%
  mutate(industry = fct_reorder(industry,change)) %>%
  ggplot(aes(x = change, y = industry, fill = race_gender)) +
  geom_col(position = "dodge") +
  scale_x_continuous(labels = percent) +
  labs(title = "Industries that suffered in employment drop in 2020",
       x = "Shift in number employed from 2019 to 2020 by gender",
       y = "",
       fill = "Gender")
```



Estimate the effect of pandemic on employment for each industry on different races.

```
compare_2019_2020 %>%
  filter(dimension == "Race") %>%
  mutate(industry = fct_reorder(industry,change)) %>%
  ggplot(aes(x = change, y = industry, fill = race_gender)) +
  geom_col(position = "dodge") +
  scale_x_continuous(labels = percent) +
  labs(title = "Industries that suffered in employment drop in 2020",
       x = "Shift in number employed from 2019 to 2020 by gender",
       y = "",
       fill = "Race")
```