San Diego Small Businesses: Data Journey

This markdown file is to allow readers of the “SB EDD White Paper v2.pdf” to recreate the data. It is the process that was used to clean the data, to prep the data and to analyze the data through graphs and charts. This markdown file is not the analysis.

As mentioned in the README document, there is an excel spreadsheet, sand$sf2018.xlsx, that comes directly from the State of California’s Department of Economic Development website. https://www.labormarketinfo.edd.ca.gov/LMID/Size_of_Business_Data.html (https://www.labormarketinfo.edd.ca.gov/LMID/Size_of_Business_Data.html)

Eighteen spreadsheets, spanning from 2001 to 2018, were manually downloaded and reformatted. Each file was imported into R, dates were added and the data sets were merged. The resulting file, raw_data.csv, is included in the GitHub repository.

The raw data set contains three performance metrics - Number of Businesses, Number of Employees, and Third Quarter Payroll - by the Size Category and the Industry. Beginning in 2017, the report switched over to using Average Monthly Employment (AME) instead of third-month employment.

```
library(tidyr)
library(dplyr)
library(tidyverse)
library(reshape2)
EDCData <- read_csv("raw_data.csv")
```

The excel spreadsheets from the EDD have some slight variations in spellings and in groupings over the years. For example “Construction & Mining” were separate industries for a few years before being put back together. To help with analysis, industry categories where given better column names.

```
EDCData2 <- EDCData %>%
  mutate(IndCategory =
      ifelse(Industry == "Transportation and Warehousing",
             "Transportation & Warehousing", ###
      ifelse(Industry == "Transportation And Warehousing",
             "Transportation & Warehousing", ###
      ifelse(Industry == "Real Estate & Rental And Leasing",
             "Real Estate & Rental/Leasing", ###
      ifelse(Industry == "Real Estate and Rental and Leasing",
             "Real Estate & Rental/Leasing", ###
      ifelse(Industry == "Finance and Insurance",
             "Finance & Insurance", ###
      ifelse(Industry == "Finance And Insurance",
             "Finance & Insurance", ###
      ifelse(Industry == "Ag, Forestry, Fishing, Hunting",
             "Agriculture", ###
      ifelse(Industry == "Construction and Mining",
             "Construction & Mining", ###
      ifelse(Industry == "Construction",
             "Construction & Mining", ###
      ifelse(Industry == "Mining",
             "Construction & Mining", ###
  Industry))))))))))) %>%
  mutate(EDDCat =
      ifelse(Data=="Payroll (in thousands)",
             "Payroll", ###
      ifelse(Data=="No. of Businesses",
             "Num_Business", ###
      ifelse(Data=="No. of Employees",
             "Num_Employees", ###
  Data)))) %>%
  select(-c(Industry, Data))
```

More cleanup of names, changing the column name from “0-4”" to “E0_4”.

```
names(EDCData2) <- c("Total", "E0_4", "E5_9","E10_19", "E20_49", "E50_99",
                     "E100_249", "E250_499", "E500_999", "E1000", "Year",
                     "IndCategory", "EDDCat"
                    )
```

Two new variables are created, E0_49, which sums across businesses with less than 50 employees, and E50P, the sum for businesses with 50 or more employees. As mentioned in the paper, adding E0_49 and E50P will not always equal Total because the EDD will not show data if there is a chance that the company could be identified.

```
EDCData2 <- EDCData2 %>%
  rowwise() %>%
  mutate(E0_49 = sum(E0_4, E5_9, E10_19, E20_49, na.rm = TRUE)) %>%
  mutate(E50P = sum(E50_99, E100_249, E250_499,
                    E500_999, E1000, na.rm = TRUE))
```

Finally, with the data prepped, a Tidy date set is created to build graphs and analysis.

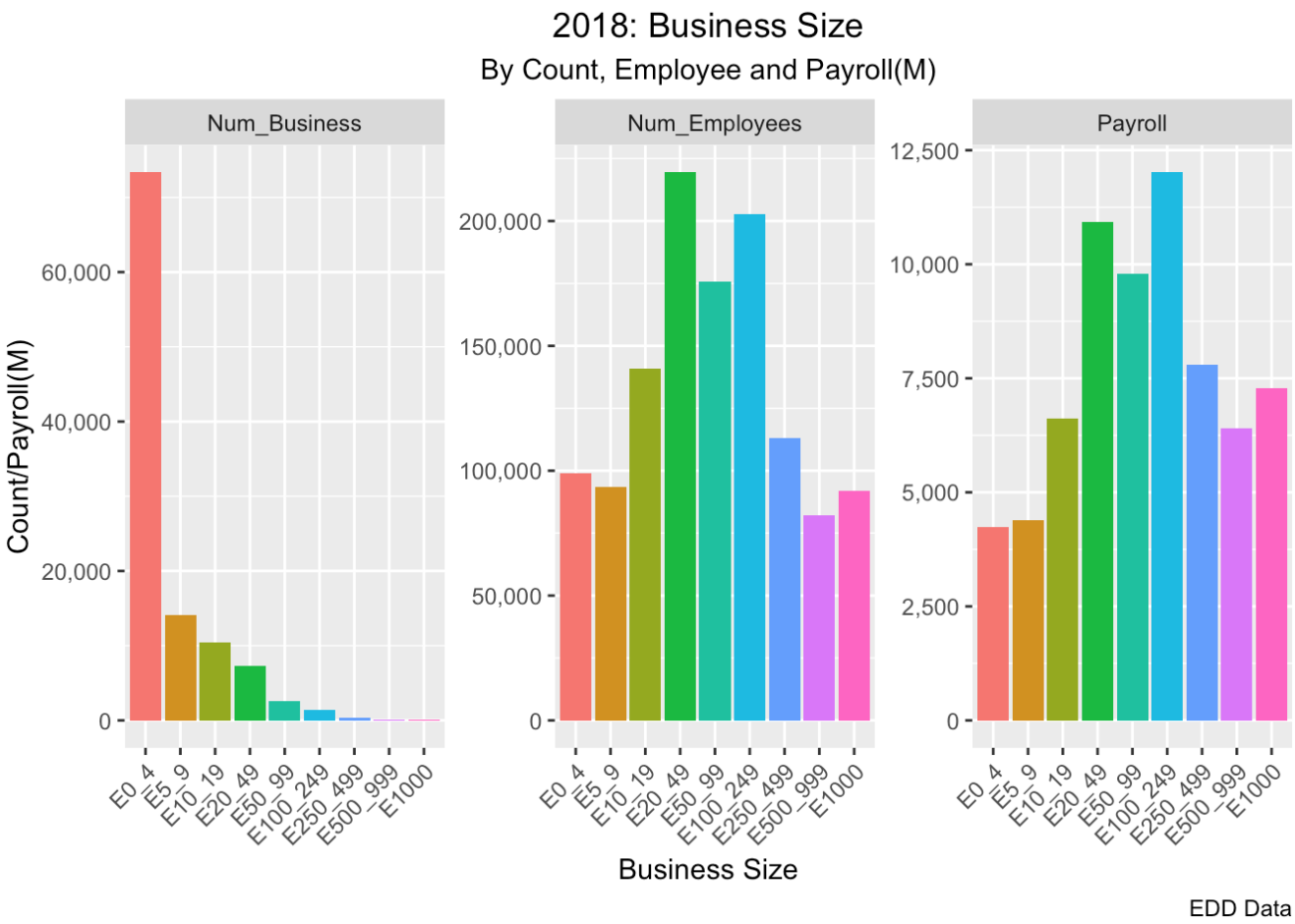
```
EDCT <- melt(
  data = EDCData2,
  id = c("Year", "IndCategory", "EDDCat"),
  variable.name = "BusinessSize",
  value.name = "Value")
```

One more data cleanup before we start exploring, summarize rows to handle the issues with construction and mining.

```
SumData <- aggregate(EDCT$Value,
           by=list(EDCT$Year, EDCT$IndCategory,
                   EDCT$EDDCat, EDCT$BusinessSize),
           FUN=sum, na.rm=TRUE)
names(SumData) <- c("Year", "IndCategory", "EDDCat",
                    "BusinessSize", "Value")
```

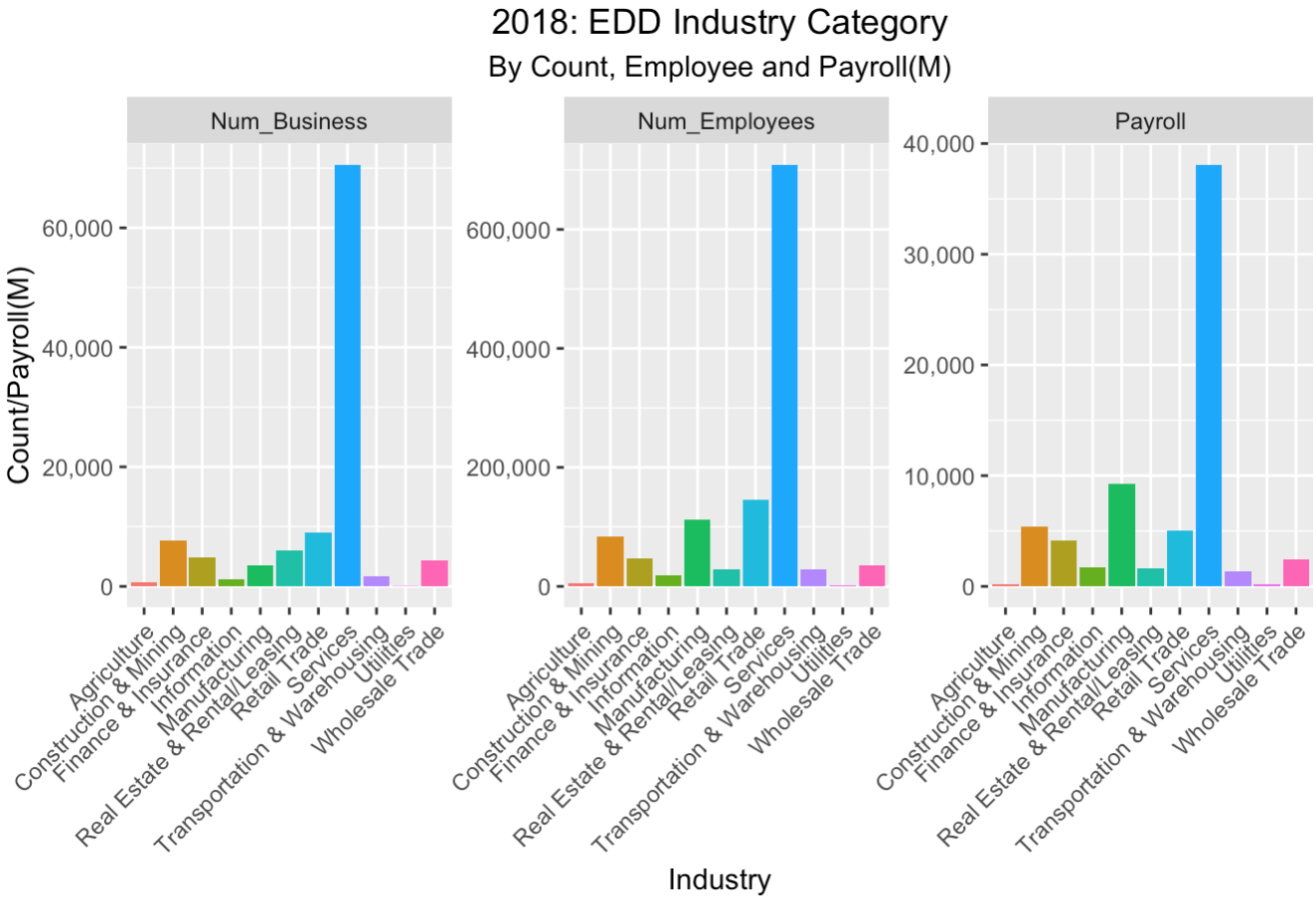
The code below allows us to see summary data of our 3 variables by Business Size. Because the data is only from the third quarter, payroll data has to be multiplied by 4. The EDD also provides payroll “in the thousands”. This still is a large number to plot, so the payroll data is presented in millions.

```
SumData %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(!BusinessSize %in% c("Total", "test", "E0_49", "E50P")) %>%
  mutate(tempV =
     ifelse(EDDCat == "Payroll", (Value*4*1000)/1000000,
     Value)) %>%
  ggplot(aes(x=BusinessSize)) +
  geom_bar(aes(weight = tempV, fill=BusinessSize)) +
  labs(title="2018: Business Size",
       subtitle="By Count, Employee and Payroll(M)",
       y="Count/Payroll(M)",
       x="Business Size",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  facet_wrap(~EDDCat, scales = "free_y")
```



Taking a look at our performance metrics by Industry Category helps us understand the biggest driver in San Diego - Services.

```
SumData %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(!BusinessSize %in% c("Total", "test", "E0_49", "E50P")) %>%
  mutate(tempV =
           ifelse(EDDCat == "Payroll", (Value*4*1000)/1000000,
                  Value)) %>%
  ggplot(aes(x=IndCategory)) +
  geom_bar(aes(weight = tempV, fill=IndCategory)) +
  labs(title="2018: EDD Industry Category",
       subtitle="By Count, Employee and Payroll(M)",
       y="Count/Payroll(M)",
       x="Industry",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  facet_wrap(~EDDCat, scales = "free_y")
```

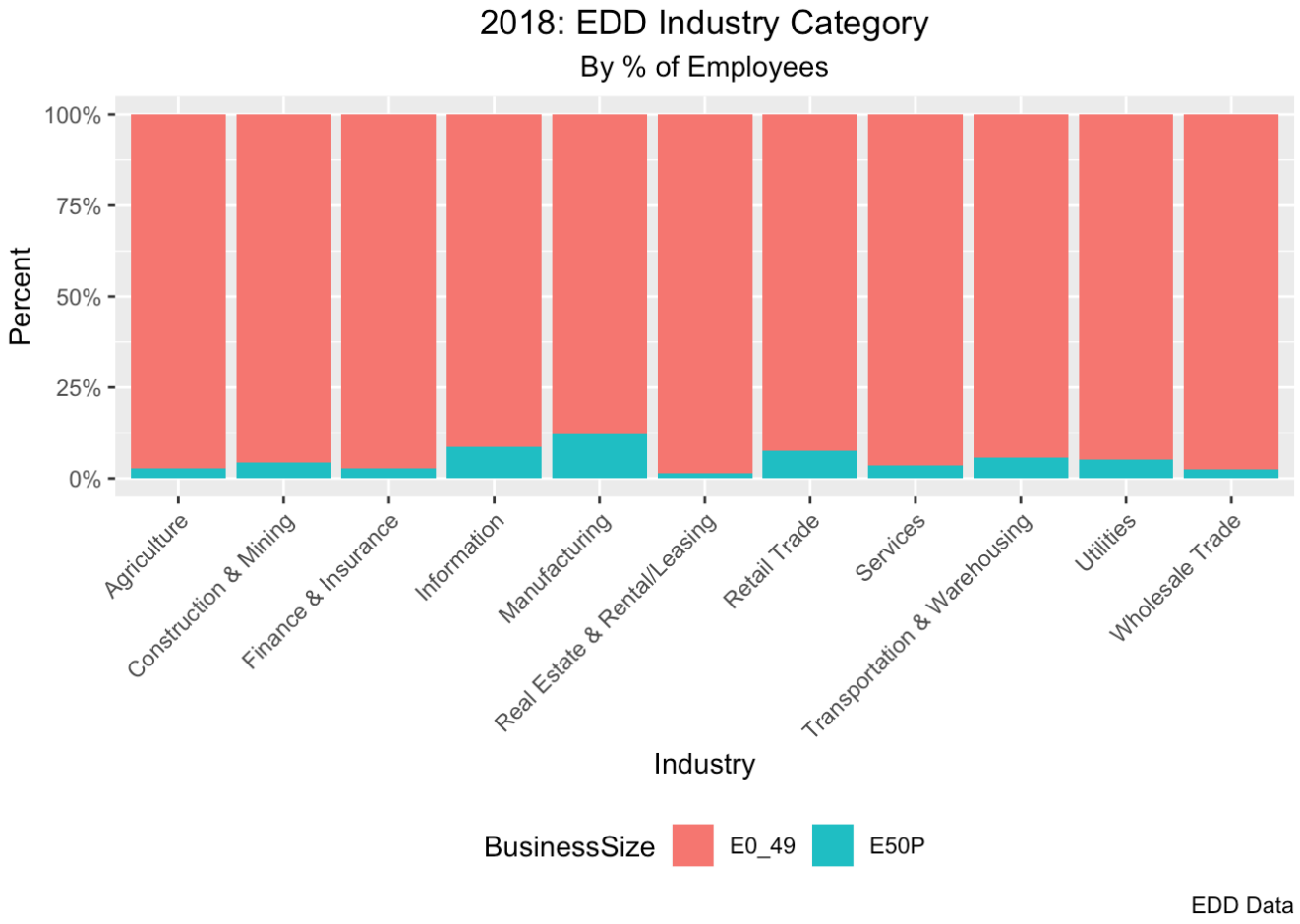


EDD Data

The goal of the analysis was to better understand the impact Small Businesses - businesses with 0-49 employees - have in San Diego. Plotting the 2 variables we created, E0_49 and E50P, as percentages helps better visualize the impact of business size.

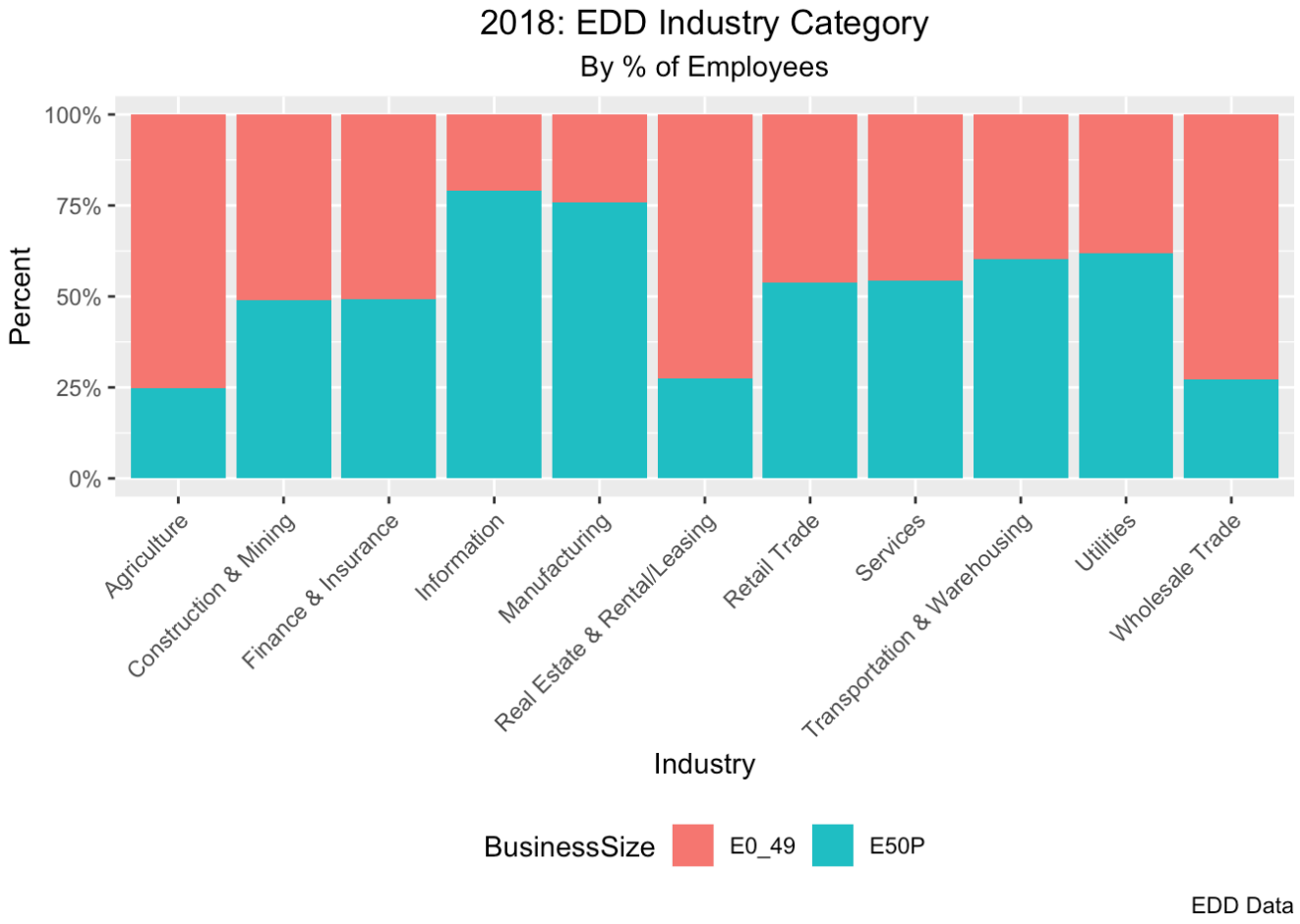
Small Businesses represent 96% of all businesses in San Diego, which is clearly represented by the graph below.

```
SumData %>%
  filter(EDDCat == "Num_Business") %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  mutate(tempV =
           ifelse(EDDCat == "Payroll", (Value*4*1000)/1000000,
                  Value)) %>%
  ggplot(aes(fill=BusinessSize, x=IndCategory, y=tempV)) +
  geom_bar(position="fill", stat="identity") +
  labs(title="2018: EDD Industry Category",
       subtitle="By % of Employees",
       y="Percent",
       x="Industry",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



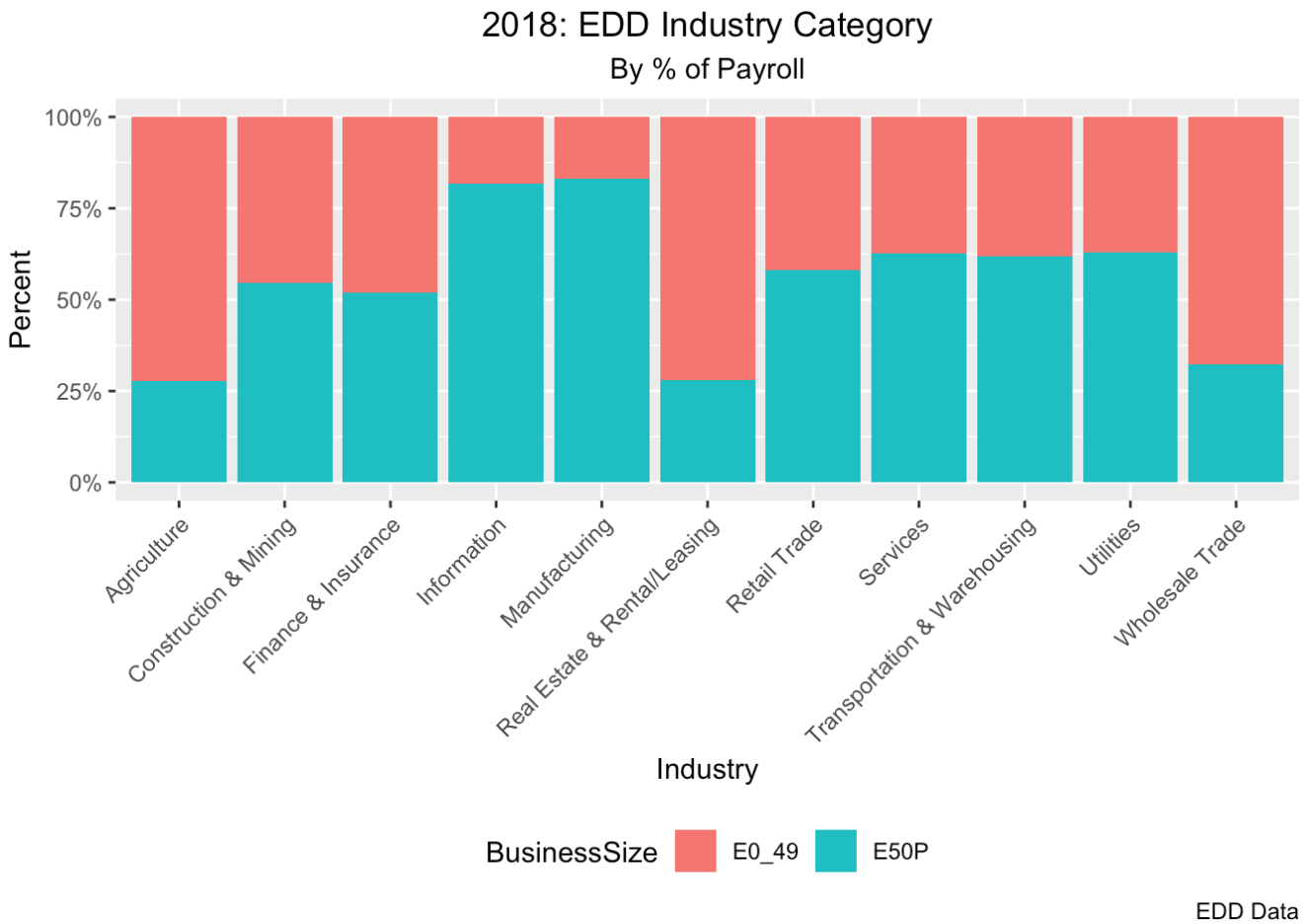
Looking at the number of employees by Industry Category, we get a better sense of where Small Business employ a bigger % of the workforce - Agriculture, Real Estate & Rental/Leasing and Wholesale Trade.

```
SumData %>%
  filter(EDDCat == "Num_Employees") %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  mutate(tempV =
           ifelse(EDDCat == "Payroll", (Value*4*1000)/1000000,
                  Value)) %>%
  ggplot(aes(fill=BusinessSize, x=IndCategory, y=tempV)) +
  geom_bar(position="fill", stat="identity") +
  labs(title="2018: EDD Industry Category",
       subtitle="By % of Employees",
       y="Percent",
       x="Industry",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



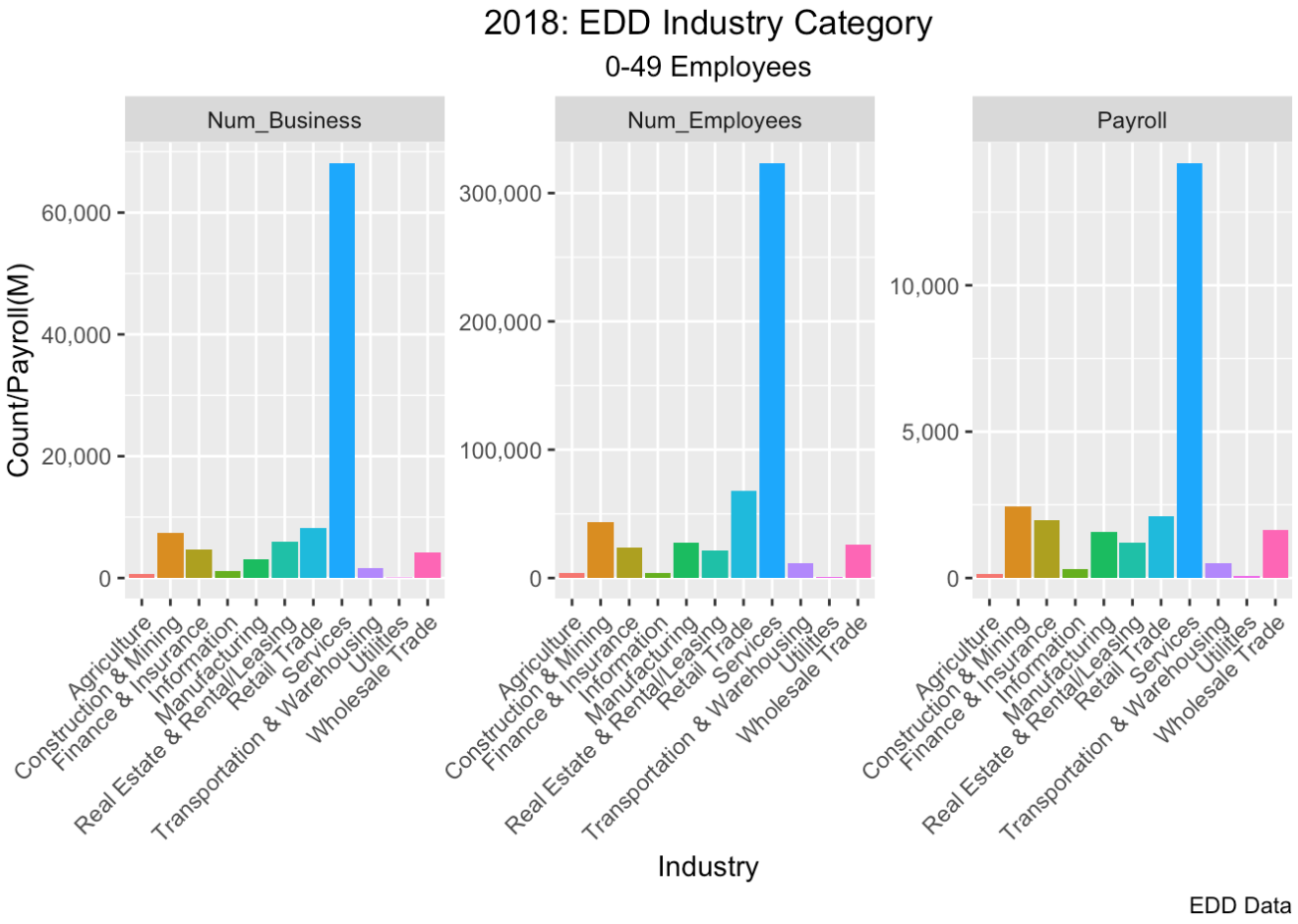
Looking at Payroll numbers, we see the same distribution as we saw with Number of Employees.

```
SumData %>%
  filter(EDDCat == "Payroll") %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  mutate(tempV =
           ifelse(EDDCat == "Payroll", (Value*4*1000)/1000000,
                  Value)) %>%
  ggplot(aes(fill=BusinessSize, x=IndCategory, y=tempV)) +
  geom_bar(position="fill", stat="identity") +
  labs(title="2018: EDD Industry Category",
       subtitle="By % of Payroll",
       y="Percent",
       x="Industry",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



However, if we look at just Small Businesses, Agriculture, Real Estate & Rental/Leasing and Wholesale Trade are relatively small industries compared to the Service sector.

```
SumData %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(BusinessSize %in% c("E0_49")) %>%
  mutate(tempV =
           ifelse(EDDCat == "Payroll", (Value*4*1000)/1000000,
                  Value)) %>%
  ggplot(aes(x=IndCategory)) +
  geom_bar(aes(weight = tempV, fill=IndCategory)) +
  labs(title="2018: EDD Industry Category",
       subtitle="0-49 Employees",
       y="Count/Payroll(M)",
       x="Industry",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  facet_wrap(~EDDCat, scales = "free_y")
```



One variable that hasn't been examined is average yearly employee pay. To create this variable given our Tidy data set, we need to recast our data.

```
Rdata <-  dcast(SumData,
                Year + BusinessSize + IndCategory ~ EDDCat)
```

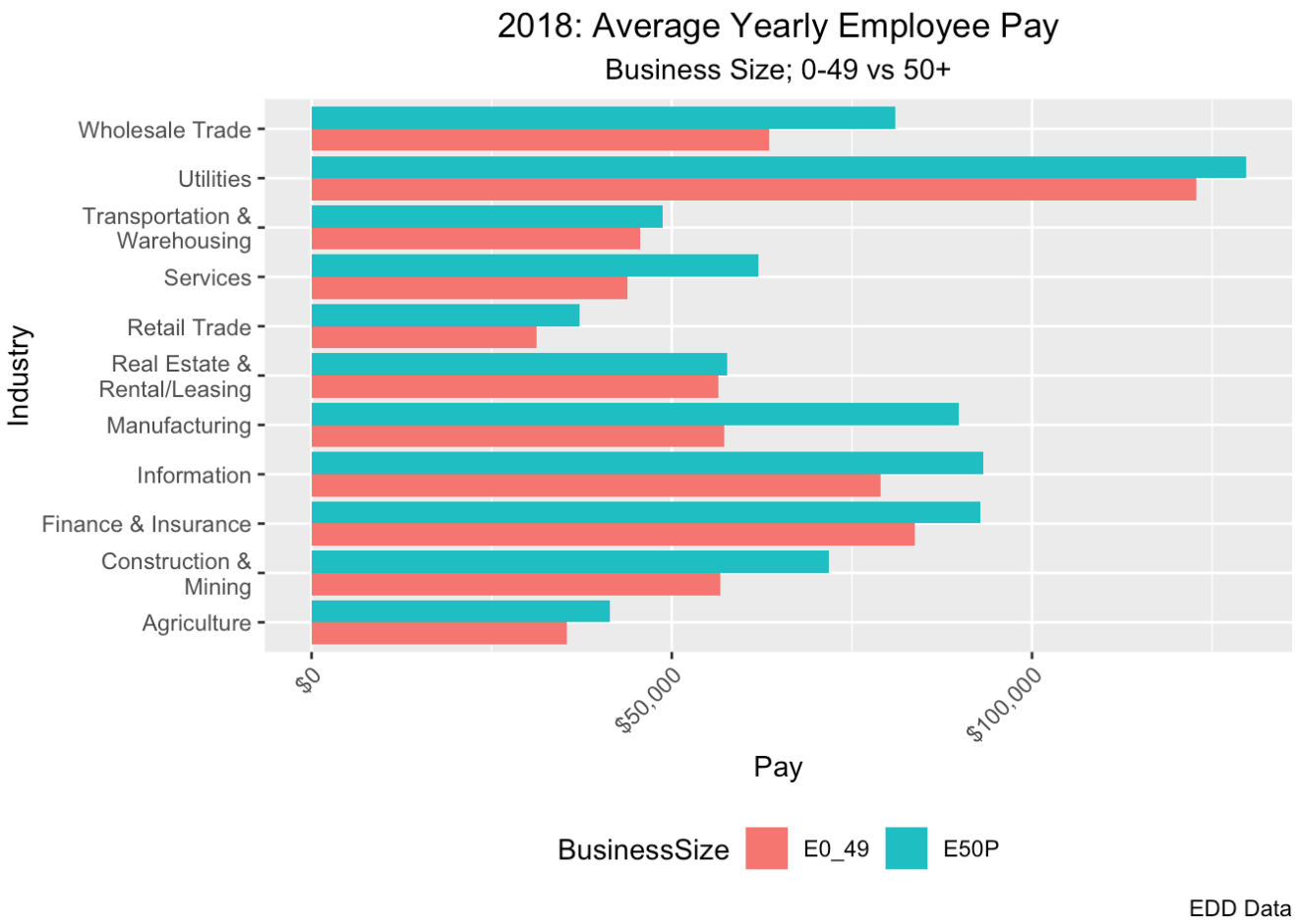
```
## Using Value as value column: use value.var to override.
```

```
Rdata <- Rdata %>%
  rowwise() %>%
  mutate(AvgPay_Yearly =
           (((Payroll*1000*4))/Num_Employees),na.rm = TRUE)
```

With this new data set, we can compare average yearly employee pay between Small Business and businesses with 50 or more employees.

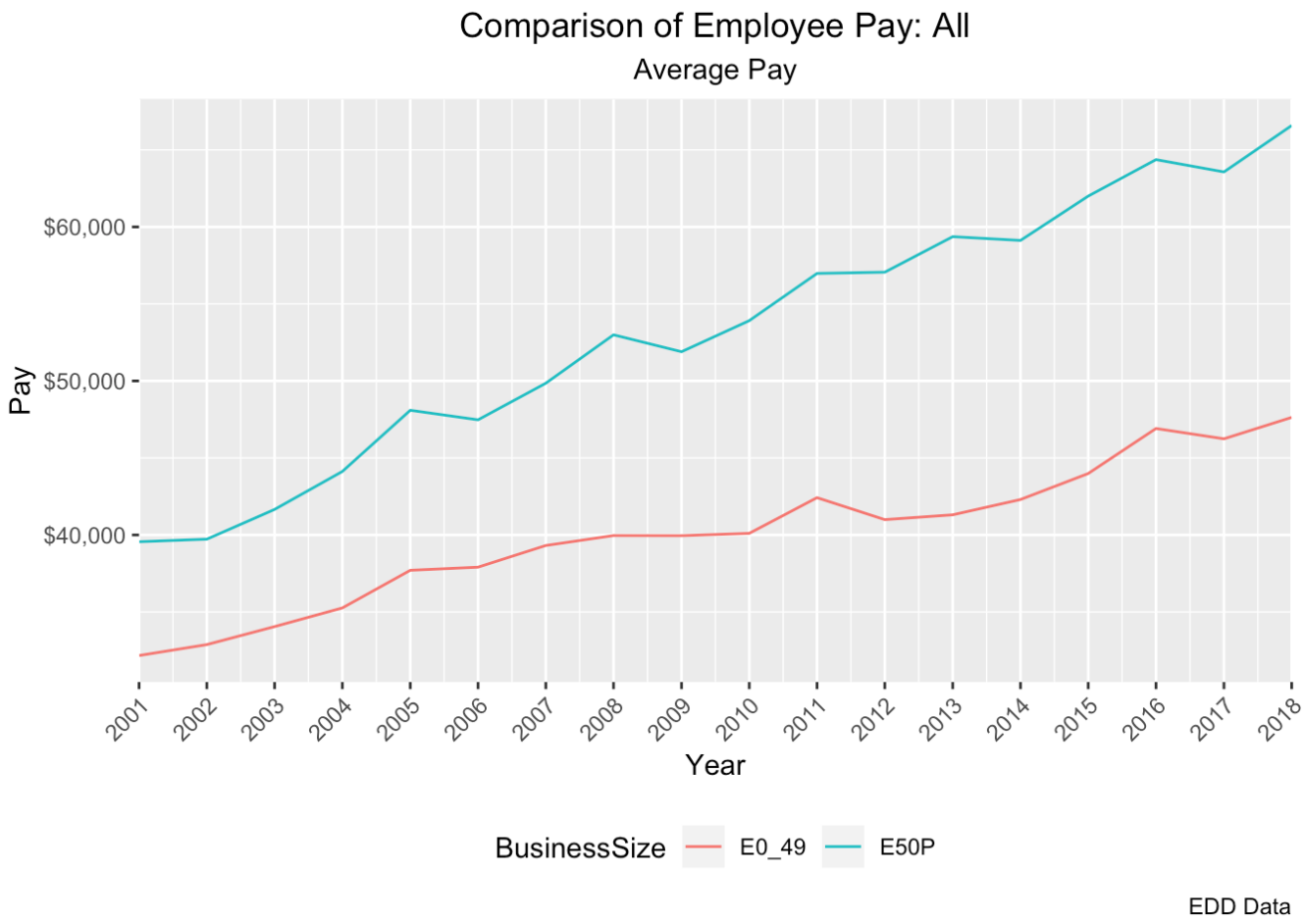
Clearly employee pay is an issue for Small Businesses.

```
Rdata %>%
  filter(Year == "2018-01-01") %>%
  filter(IndCategory != "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  ggplot(aes(x=IndCategory, y=AvgPay_Yearly, fill=BusinessSize)) +
  geom_bar(stat="identity", position=position_dodge())+
  #ggplot(aes(x=IndCategory, y = AvgPay_Yearly,
  #           group=BusinessSize, color=BusinessSize)) +
  #geom_bar() +
  labs(title="2018: Average Yearly Employee Pay",
       subtitle="Business Size; 0-49 vs 50+",
       y="Pay",
       x="Industry",
       caption="EDD Data") +
  scale_y_continuous(labels = scales::dollar) +
  scale_x_discrete(labels = function(x) str_wrap(x, width = 20)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  coord_flip()
```



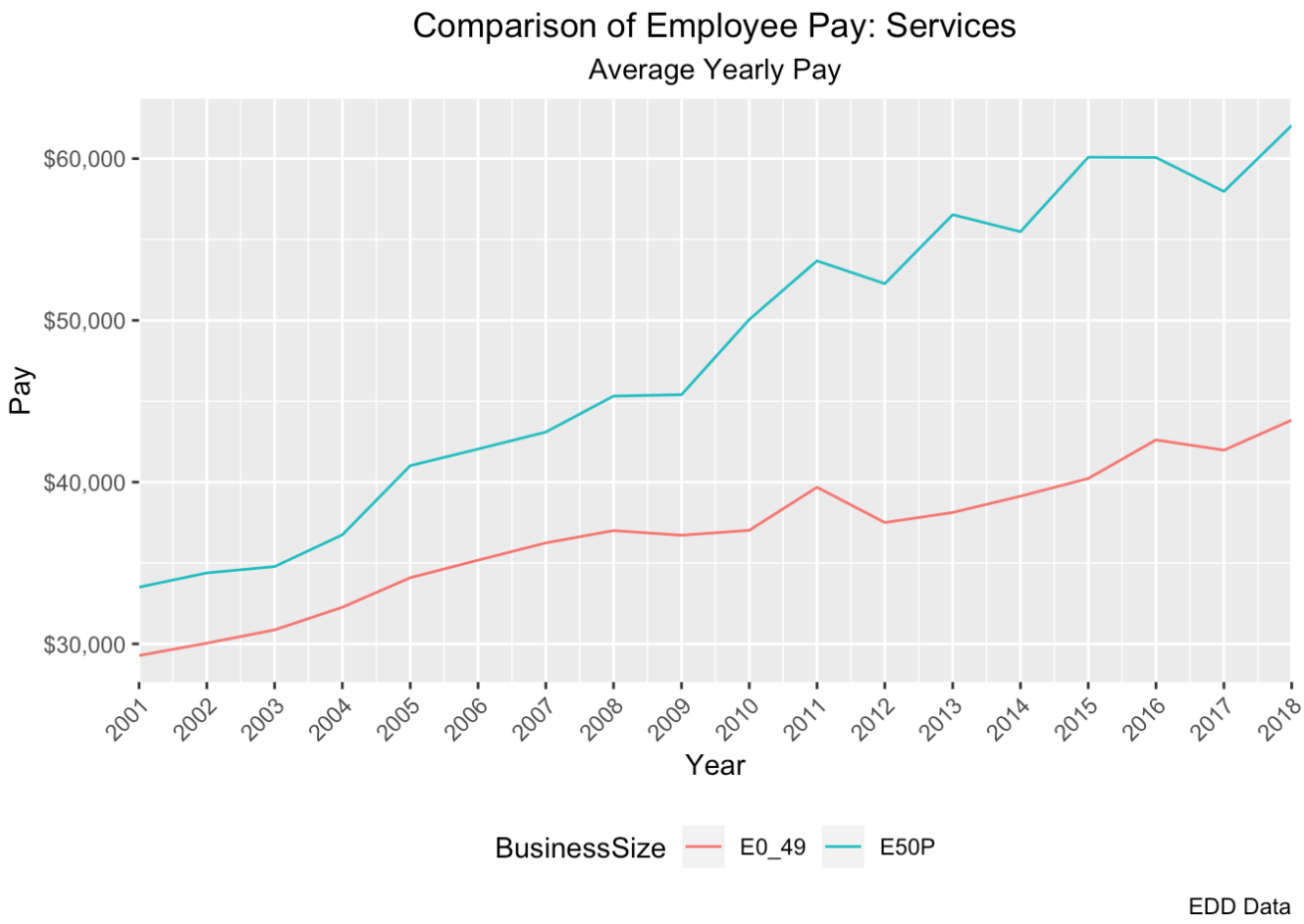
We've only been looking at data for 2018, but there is 18 yeas of data available to review. Re-plotting the bar graph above as a time series, we can see how average pay changes over time for our two Business Sizes.

```
Rdata %>%
  filter(IndCategory == "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  ggplot(aes(x=Year, y = AvgPay_Yearly,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Comparison of Employee Pay: All",
       subtitle="Average Pay",
       y="Pay",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::dollar) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



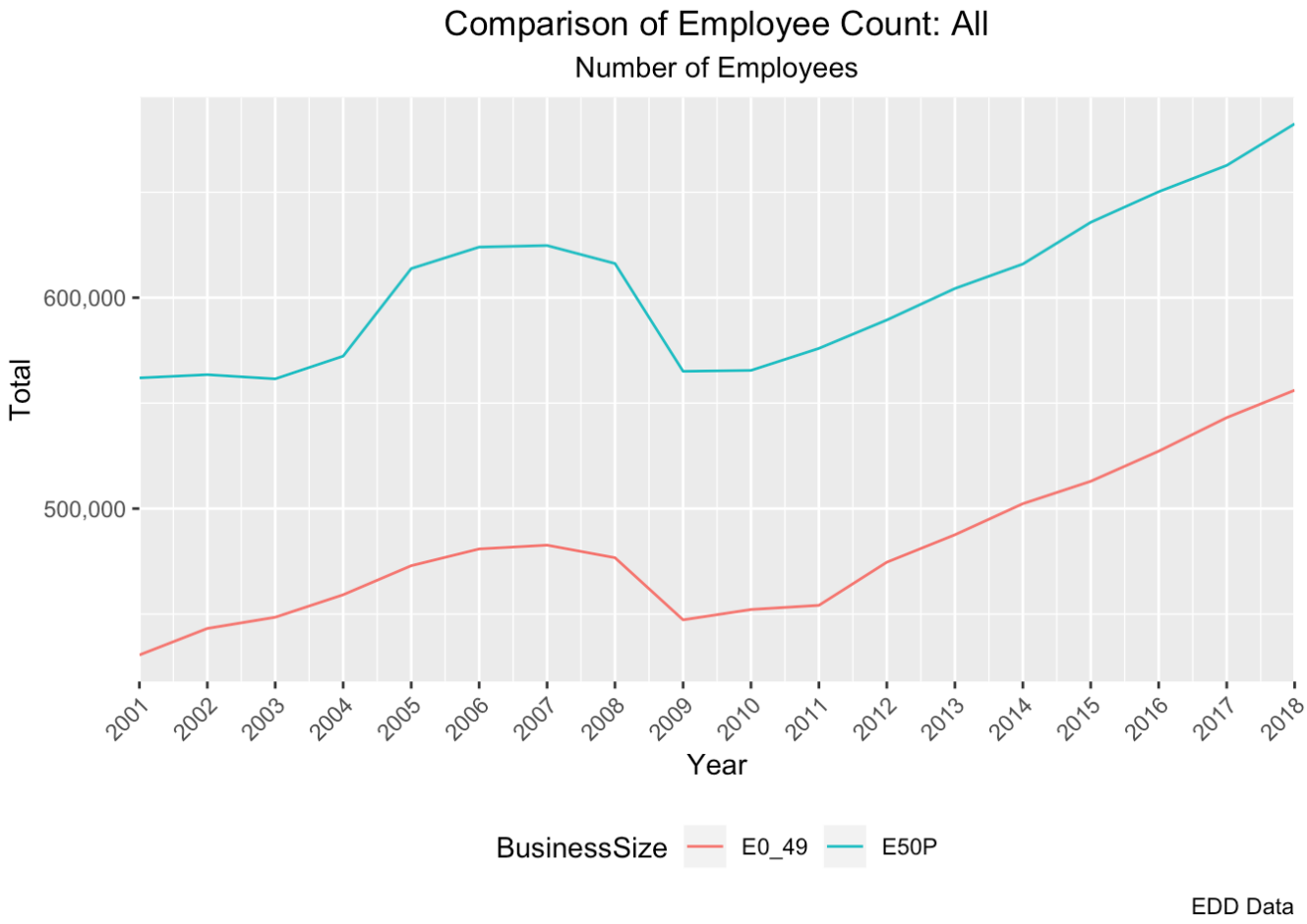
Taking a look more specifically at the Service sector.

```
Rdata %>%
  filter(IndCategory == "Services") %>%
  #filter(IndCategory == "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  ggplot(aes(x=Year, y = AvgPay_Yearly,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Comparison of Employee Pay: Services",
       subtitle="Average Yearly Pay",
       y="Pay",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::dollar) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



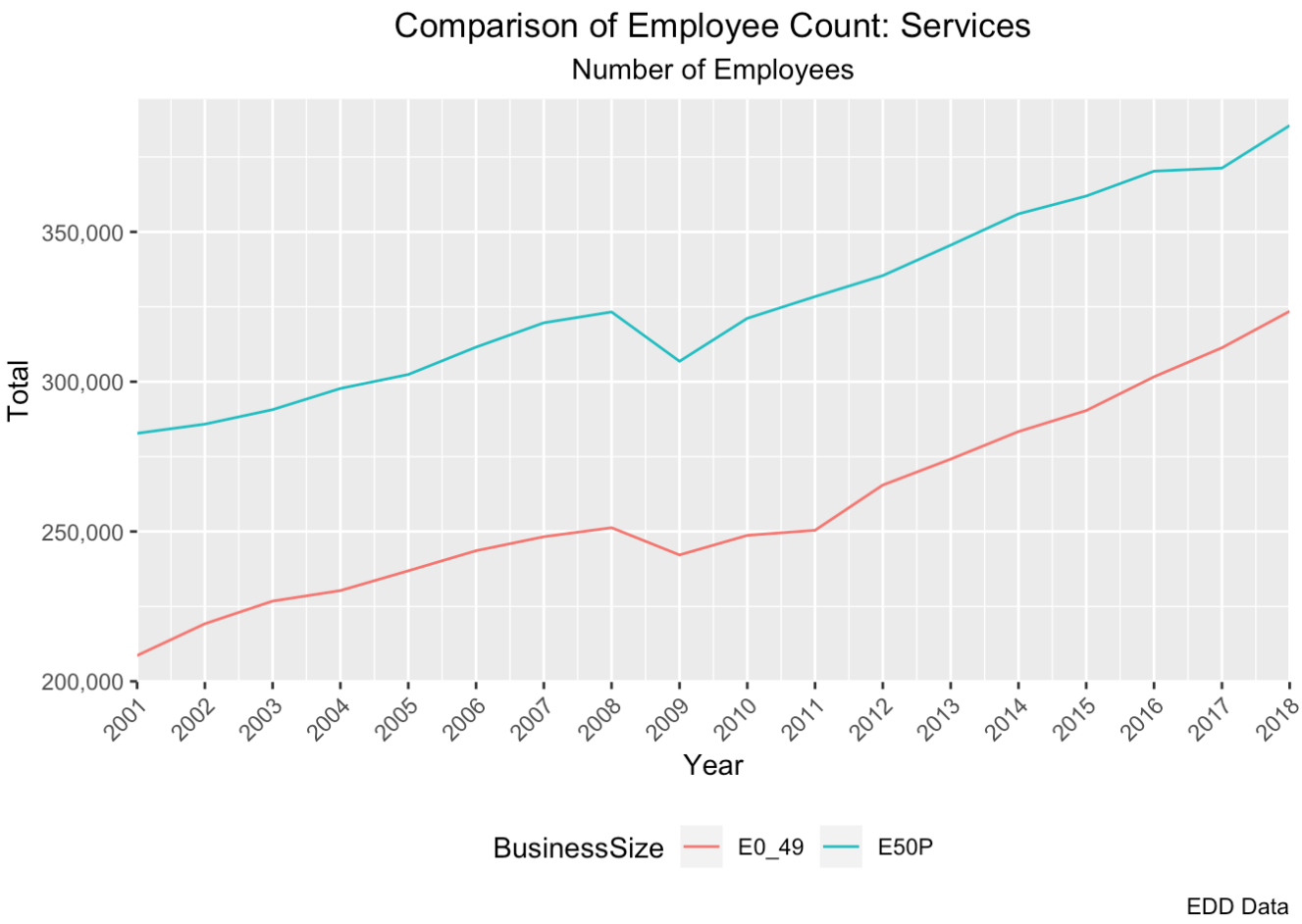
We can apply the time series plots to the other metrics we have, including Number of Employees.

```
Rdata %>%
  filter(IndCategory == "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  ggplot(aes(x=Year, y = Num_Employees,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Comparison of Employee Count: All",
       subtitle="Number of Employees",
       y="Total",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



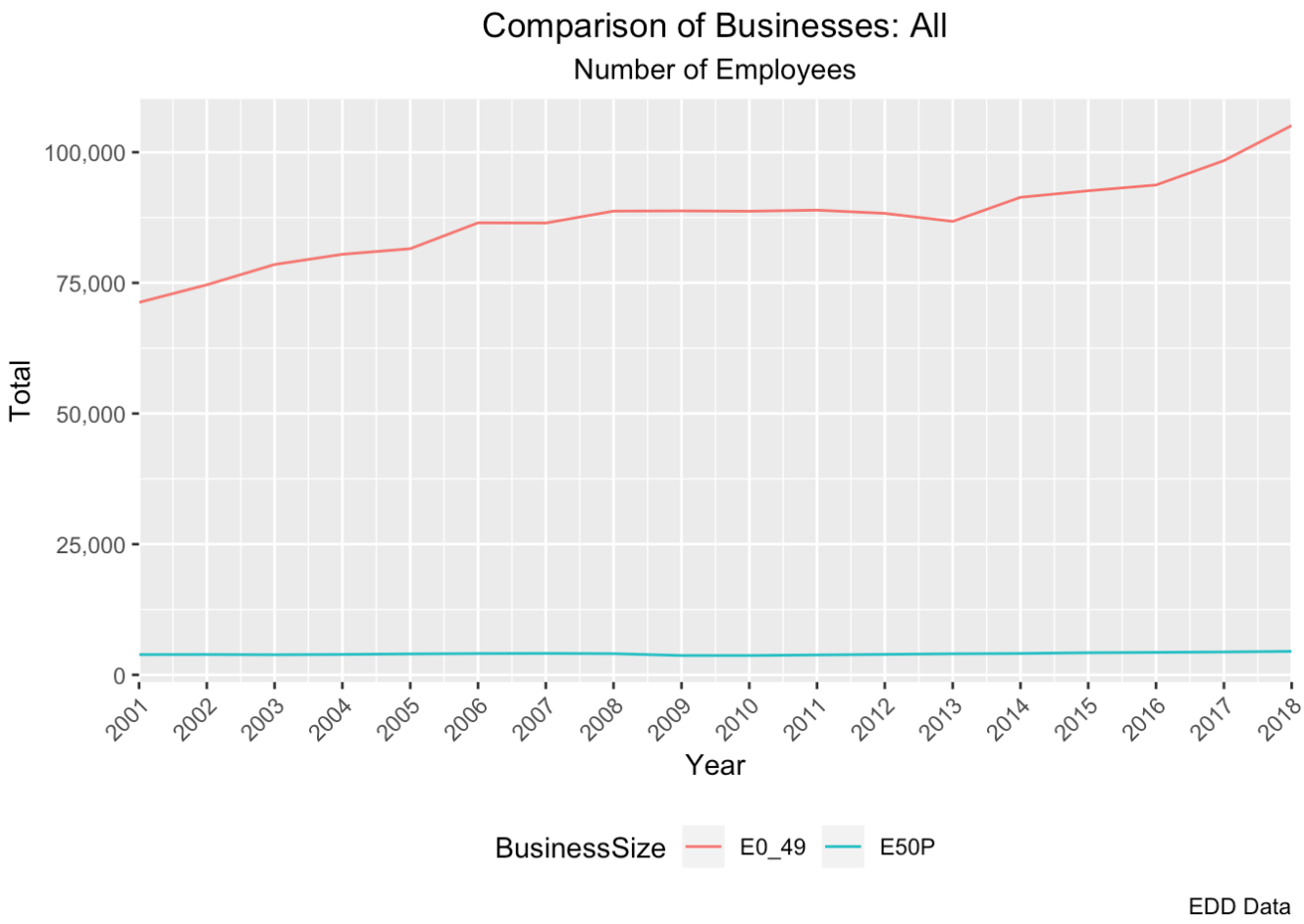
Services specifically.

```
Rdata %>%
  filter(IndCategory == "Services") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  ggplot(aes(x=Year, y = Num_Employees,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Comparison of Employee Count: Services",
       subtitle="Number of Employees",
       y="Total",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



Since 2013, there seems to have been a rapid growth in the number of Small Businesses.

```
Rdata %>%
  filter(IndCategory == "Total") %>%
  filter(BusinessSize %in% c("E0_49", "E50P")) %>%
  ggplot(aes(x=Year, y = Num_Business,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Comparison of Businesses: All",
       subtitle="Number of Employees",
       y="Total",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



Looking at percent changes is another way to evaluate the data. This next section first creates new variables that can then be plotted.

```
ts49 <- Rdata %>%
  filter(IndCategory == "Total") %>%
  filter(BusinessSize %in% c("E0_49")) %>%
  group_by(BusinessSize) %>%
  arrange(Year) %>%
  mutate(PC_APay = (AvgPay_Yearly-lag(AvgPay_Yearly))/lag(AvgPay_Yearly)) %>%
  mutate(PC_Pay = (Payroll-lag(Payroll))/lag(Payroll)) %>%
  mutate(PC_NEmp = (Num_Employees-lag(Num_Employees))/lag(Num_Employees)) %>%
  mutate(PC_NBus = (Num_Business-lag(Num_Business))/lag(Num_Business))
```

```
## Warning: Grouping rowwise data frame strips rowwise nature
```

```
ts50 <- Rdata %>%
  filter(IndCategory == "Total") %>%
  filter(BusinessSize %in% c("E50P")) %>%
  group_by(BusinessSize) %>%
  arrange(Year) %>%
  mutate(PC_APay = (AvgPay_Yearly-lag(AvgPay_Yearly))/lag(AvgPay_Yearly)) %>%
  mutate(PC_Pay = (Payroll-lag(Payroll))/lag(Payroll)) %>%
  mutate(PC_NEmp = (Num_Employees-lag(Num_Employees))/lag(Num_Employees)) %>%
  mutate(PC_NBus = (Num_Business-lag(Num_Business))/lag(Num_Business))
```

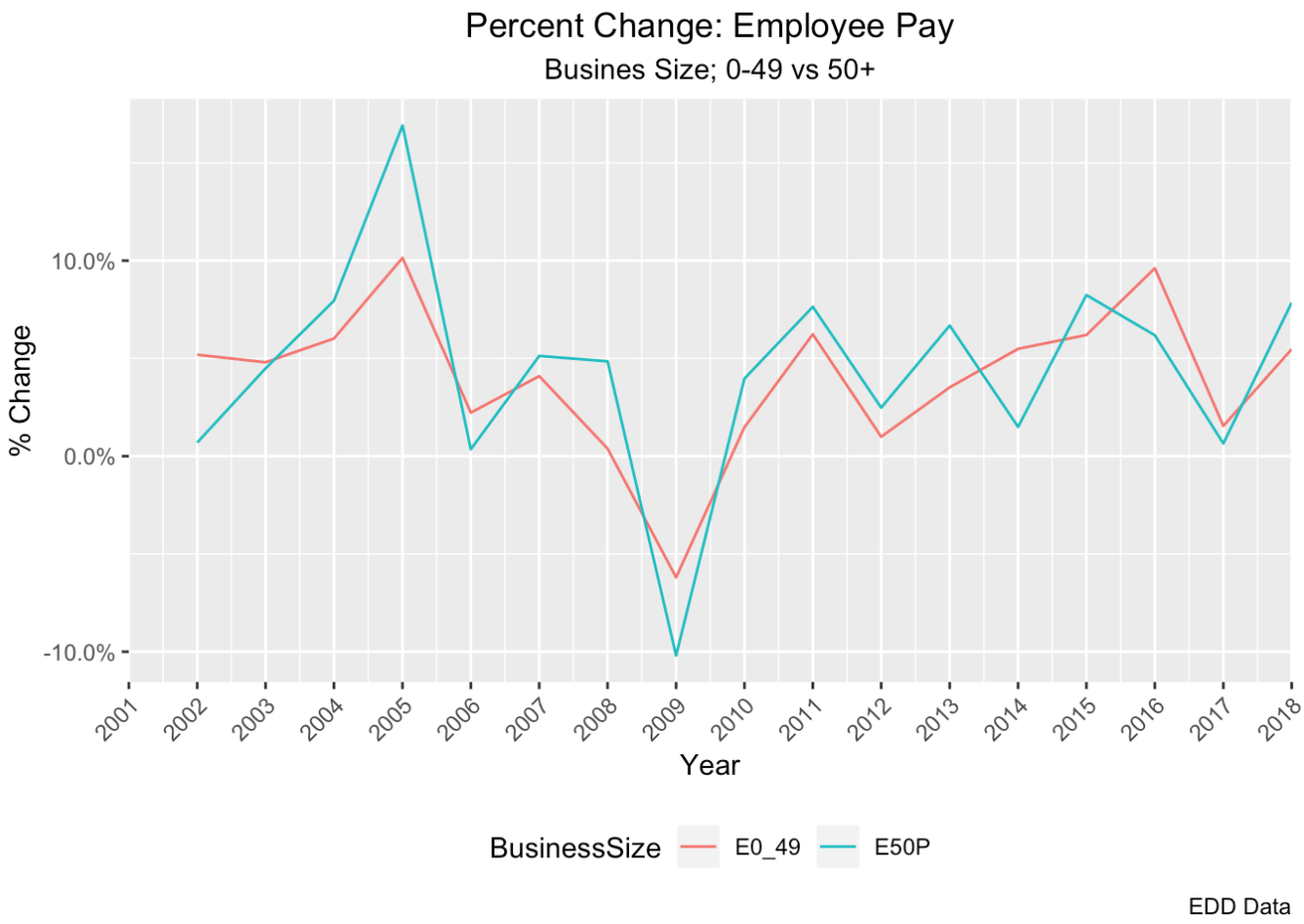
```
## Warning: Grouping rowwise data frame strips rowwise nature
```

```
tsdata <- rbind(ts49, ts50)
```

First looking at percent change for employee pay.

```
tsdata %>%
  ggplot(aes(x=Year, y = PC_Pay,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change: Employee Pay",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

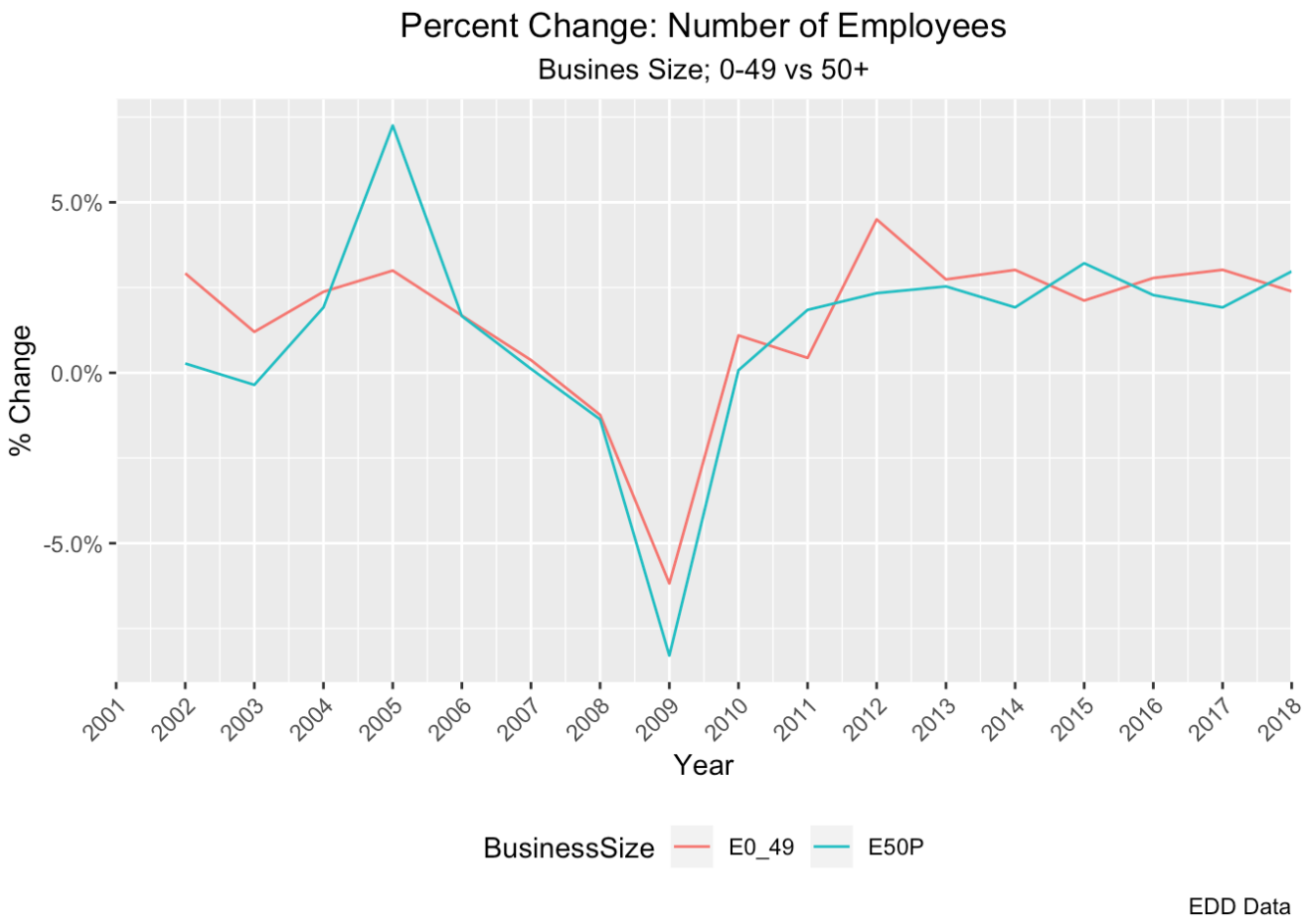
```
## Warning: Removed 2 rows containing missing values (geom_path).
```



Number of Employees

```
tsdata %>%
  ggplot(aes(x=Year, y = PC_NEmp,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change: Number of Employees",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

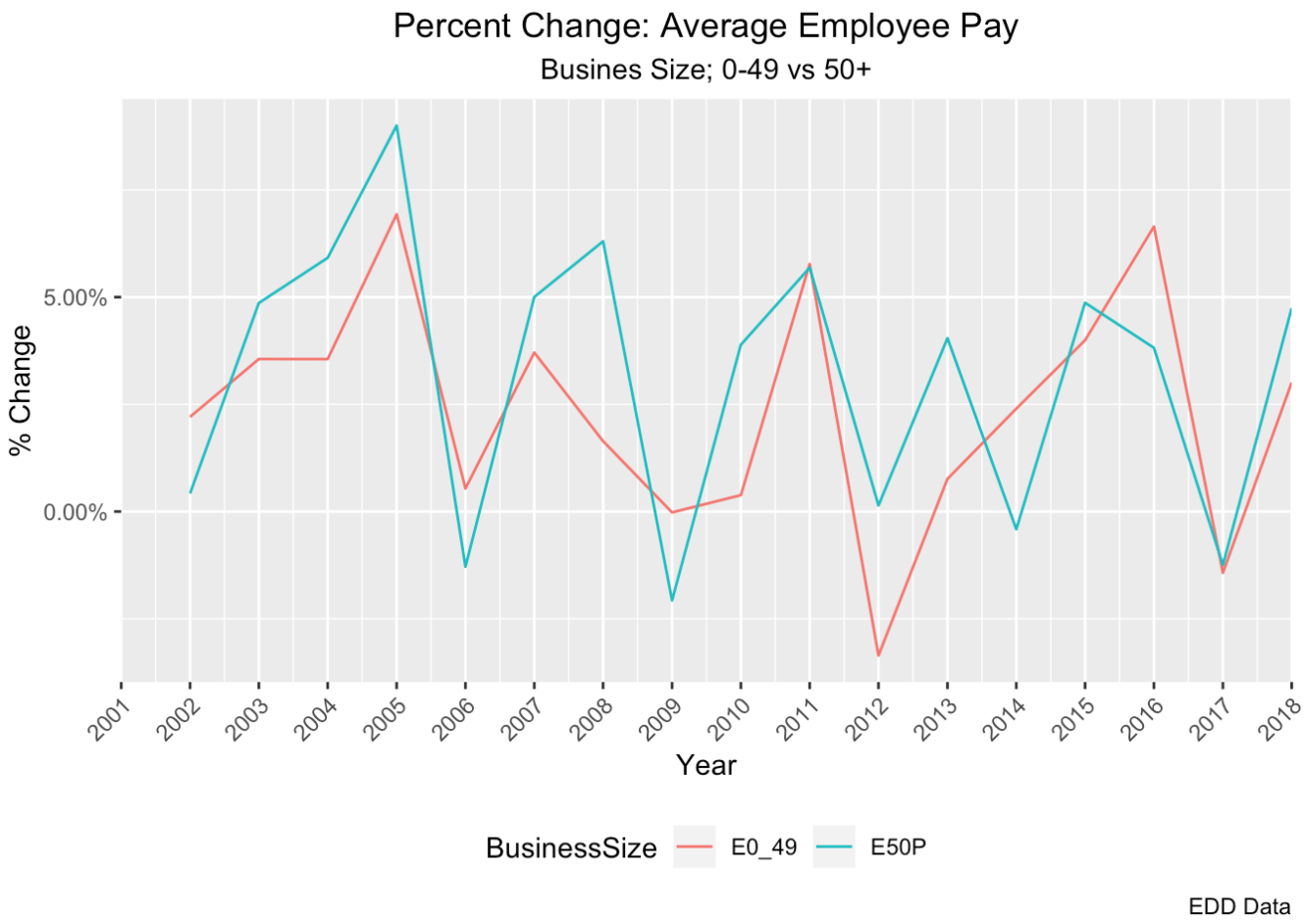
```
## Warning: Removed 2 rows containing missing values (geom_path).
```



Average Employee Pay

```
tsdata %>%
  ggplot(aes(x=Year, y = PC_APay,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change: Average Employee Pay",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

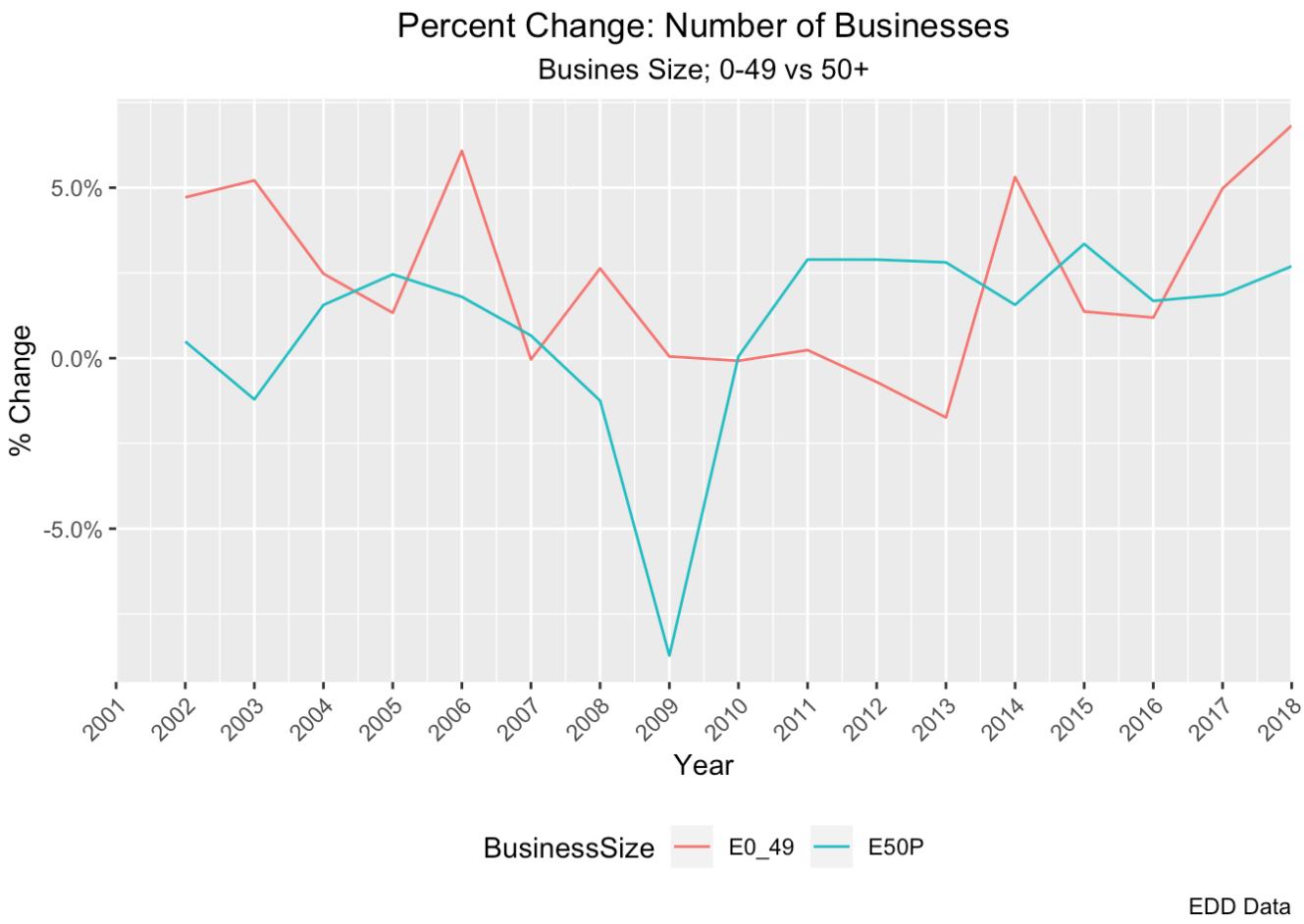
```
## Warning: Removed 2 rows containing missing values (geom_path).
```



Finally, number of businesses.

```
tsdata %>%
  ggplot(aes(x=Year, y = PC_NBus,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change: Number of Businesses",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

```
## Warning: Removed 2 rows containing missing values (geom_path).
```



In R, there can be more than one way to create variables. Here is another way to create the lag variables we are looking at. The next graphs will be specifically for Services.

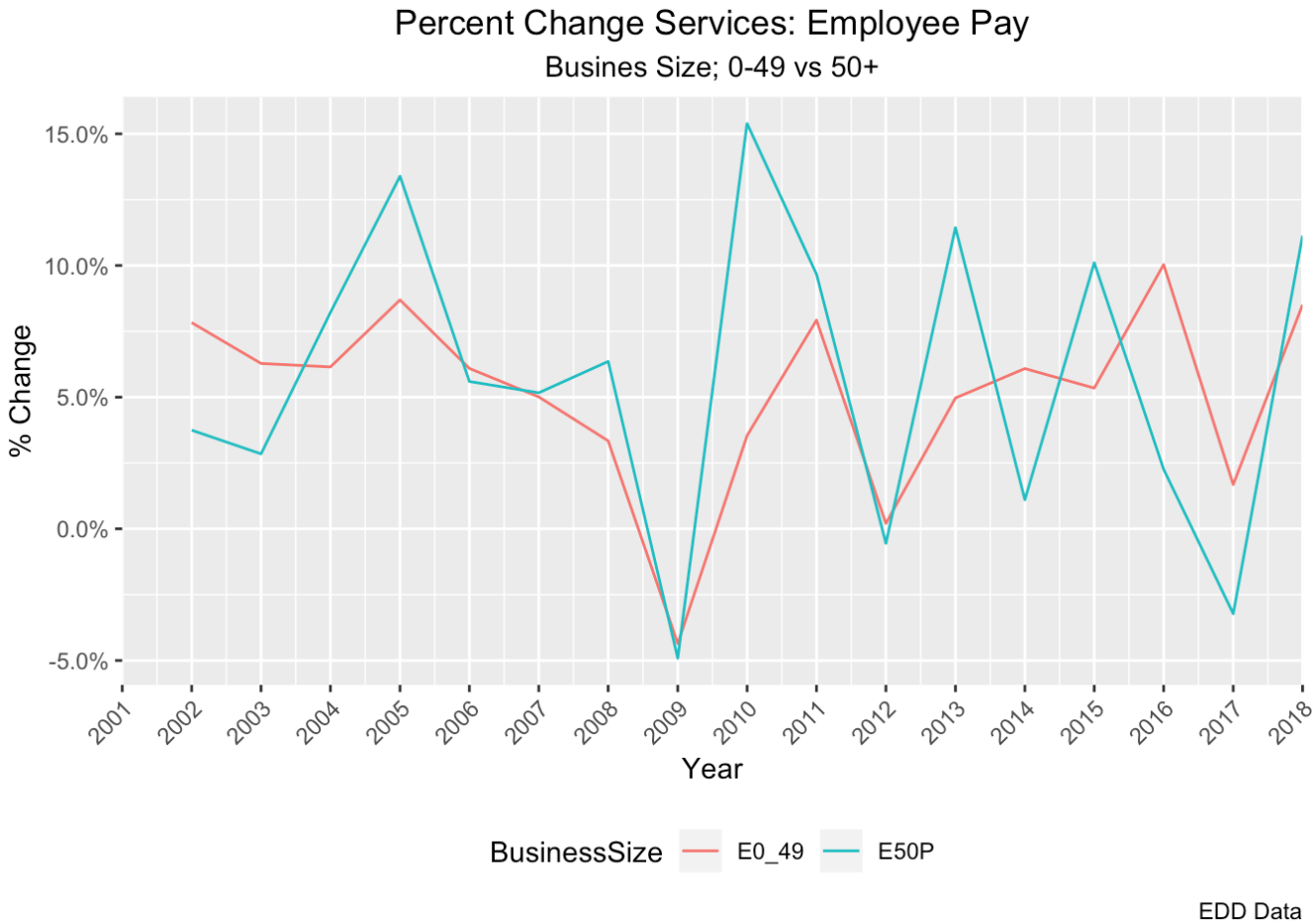
```
tt <- Rdata %>%
  filter(IndCategory == "Services") %>%
  filter(BusinessSize %in% c("E50P","E0_49")) %>%
  group_by(BusinessSize) %>%
  arrange(Year) %>%
  mutate(PC_APay = (AvgPay_Yearly-lag(AvgPay_Yearly))/lag(AvgPay_Yearly)) %>%
  mutate(PC_Pay = (Payroll-lag(Payroll))/lag(Payroll)) %>%
  mutate(PC_NEmp = (Num_Employees-lag(Num_Employees))/lag(Num_Employees)) %>%
  mutate(PC_NBus = (Num_Business-lag(Num_Business))/lag(Num_Business))
```

```
## Warning: Grouping rowwise data frame strips rowwise nature
```

Employee pay for the Service sector.

```
tt %>%
  ggplot(aes(x=Year, y = PC_Pay,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change Services: Employee Pay",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

```
## Warning: Removed 2 rows containing missing values (geom_path).
```



Change in Number of Employees for the Service sector.

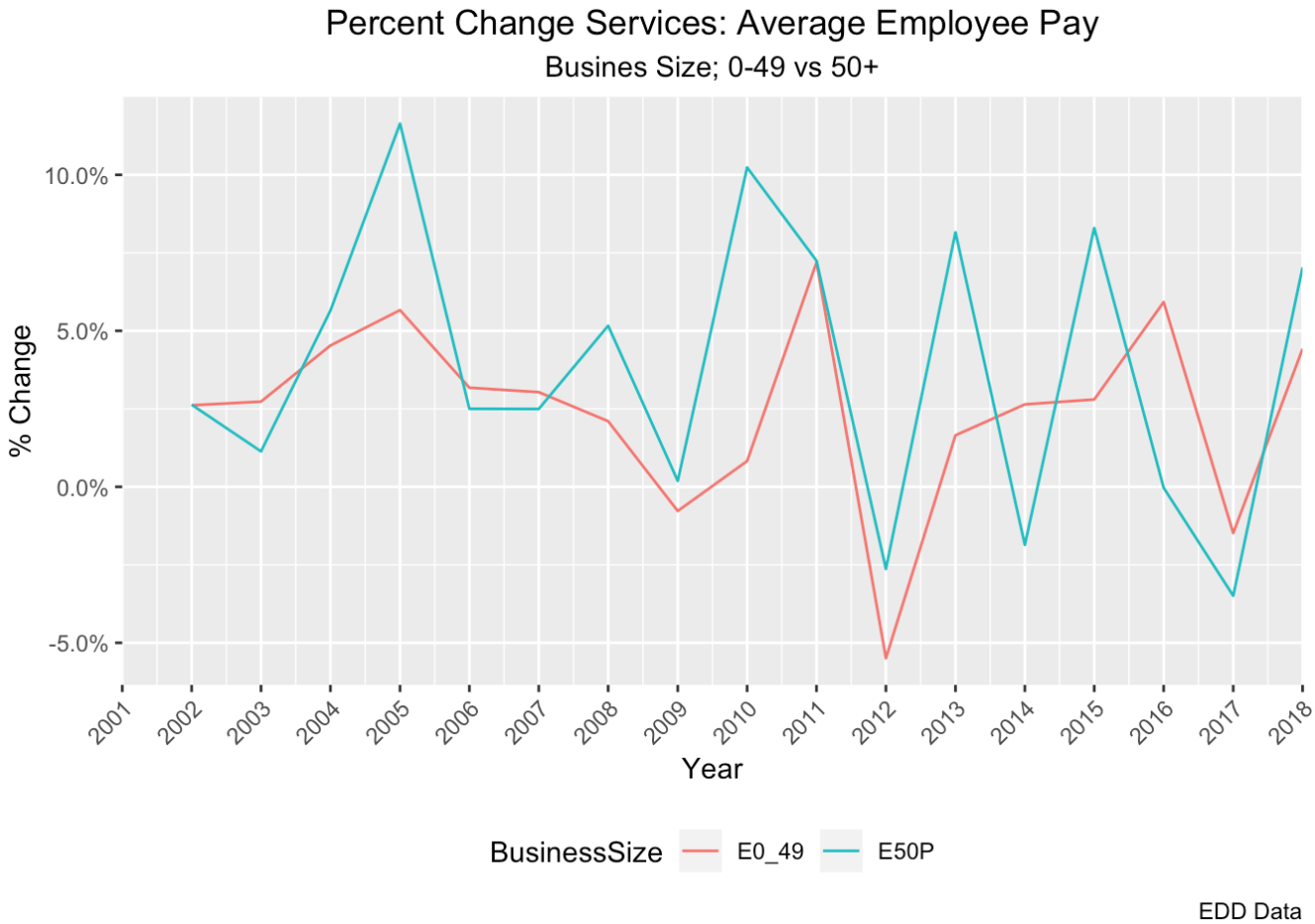
```
tt %>%
  ggplot(aes(x=Year, y = PC_NEmp,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change Services: Number of Employees",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

```
## Warning: Removed 2 rows containing missing values (geom_path).
```

Average Employee Pay for Service sector.

```
tt %>%
  ggplot(aes(x=Year, y = PC_APay,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change Services: Average Employee Pay",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

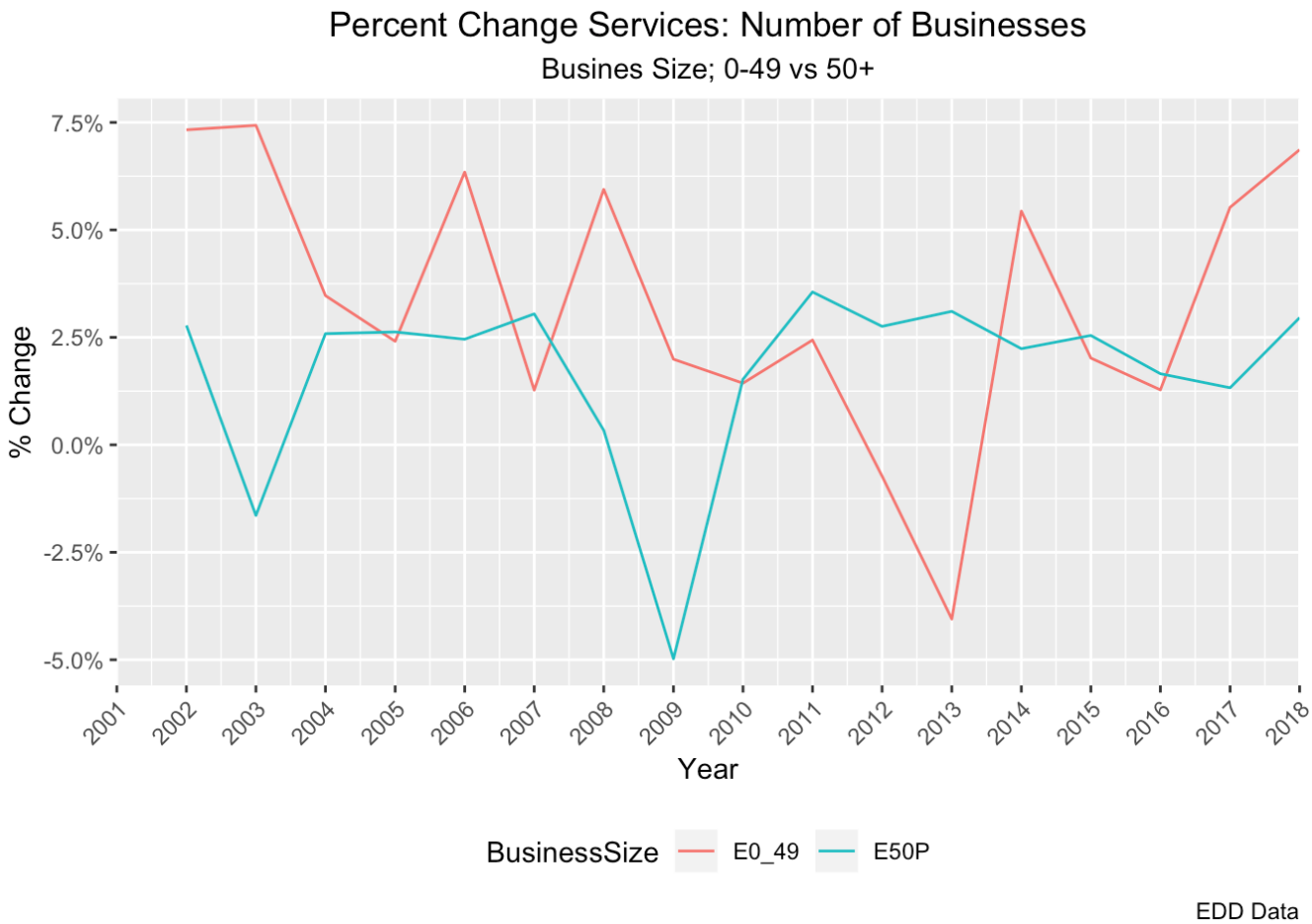
```
## Warning: Removed 2 rows containing missing values (geom_path).
```



Change in Number of Business in the Service sector.

```
tt %>%
  ggplot(aes(x=Year, y = PC_NBus,
             group=BusinessSize, color=BusinessSize)) +
  geom_line() +
  labs(title="Percent Change Services: Number of Businesses",
       subtitle="Busines Size; 0-49 vs 50+",
       y="% Change",
       x="Year",
       caption="EDD Data") +
  scale_x_date(date_labels="%Y",
               date_breaks  ="1 year",
               expand = c(0,0)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

```
## Warning: Removed 2 rows containing missing values (geom_path).
```



Overall, the data from the EDD was easy to use and insightful.