p8106_mid_project

wq2151

read data

```
# read data
stock = read_csv("./data/2018_Financial_Data.csv") %>%
  janitor::clean_names() %>%
  rename(price2019 = "x2019_price_var_percent")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

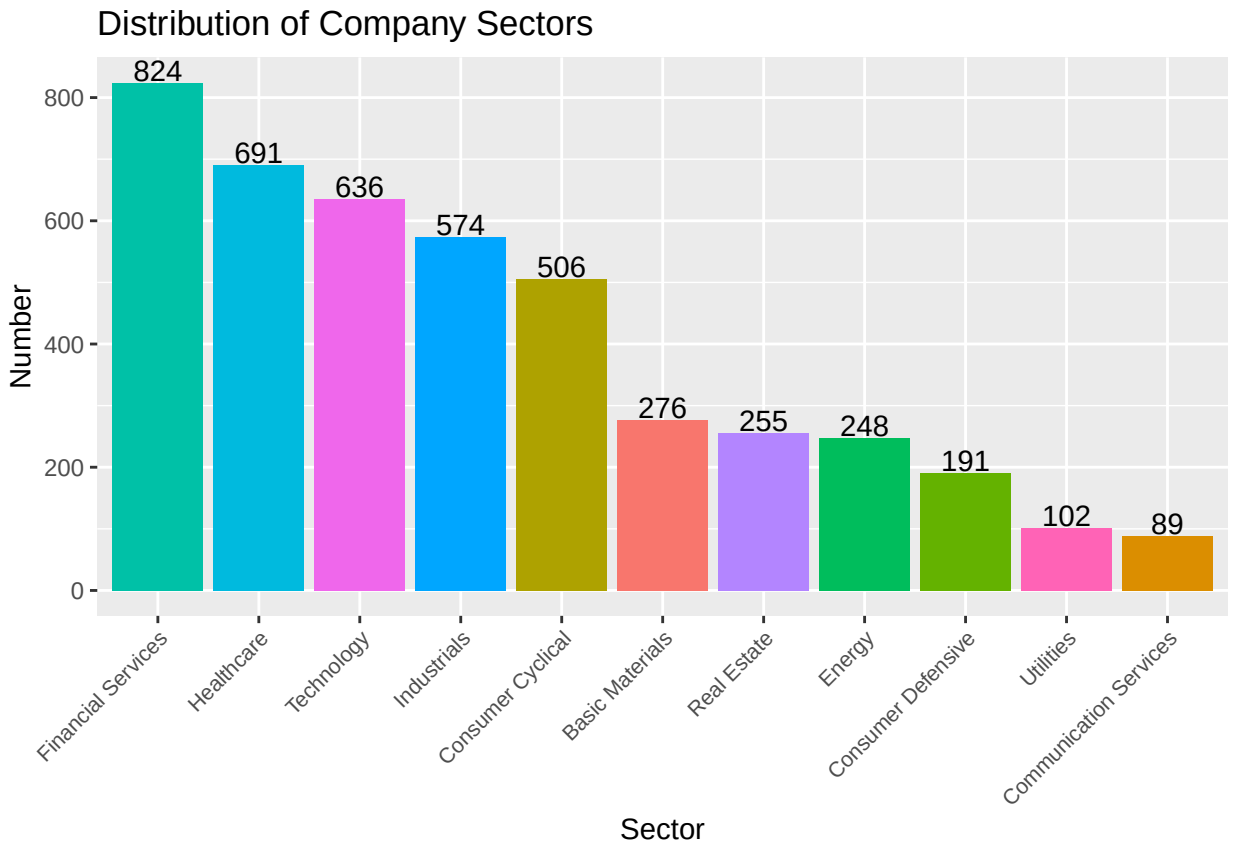
```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   X1 = col_character(),
##   Sector = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

EDA and viz

```
stock %>%
  mutate(class = as.factor(class)) %>%
  group_by(class) %>%
  summarise(number = n()) %>%
  ggplot(aes(x = class, y = number)) +
  geom_col(aes(fill = class)) +
  geom_text(aes(label = number), vjust = -0.1) +
  labs(title = "Distribution of Stock Class",
       x = "0 for BUY, 1 for SELL", y = "Number")
```


```
stock %>%
  group_by(sector) %>%
  summarise(number = n()) %>%
  ggplot(aes(x = reorder(sector, -number), y = number)) +
  geom_col(aes(fill = sector)) +
  geom_text(aes(label = number), vjust = -0.1) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        legend.position = "none") +
  labs(title = "Distribution of Company Sectors",
       x = "Sector", y = "Number")
```




```
# correlation plot (took too long and bad output)
# corrplot::corrplot(cor(stock %>% select(-x1, -sector)), method = "square", use = "pairwise.complete.o
```