Application Exercise 4

Question 1

```
library(readxl)
dja_performance_report_monthly <- read_excel("dja-performance-report-monthly.xls",
     sheet = "dja-performance-report-monthly",
     skip = 5)

class(dja_performance_report_monthly)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
dj <- dja_performance_report_monthly
```

Question 2

```
dj_new <- (dj[c(2,3)])[complete.cases(dj[c(2,3)]),]
dj_new
```

```
## # A tibble: 1,492 x 2
##    `Effective Date` `Close Value`
##    <chr>                    <dbl>
##  1 05/29/1896                40.6
##  2 06/30/1896                36.2
##  3 07/31/1896                32.0
##  4 08/31/1896                32.0
##  5 09/30/1896                36.0
##  6 10/31/1896                39.5
##  7 11/30/1896                41
##  8 12/31/1896                40.4
##  9 01/30/1897                42.6
## 10 02/27/1897                41.7
## # ... with 1,482 more rows
```

Question 3

```
dj_new <- dj_new %>%
  rename(
    date = 'Effective Date',
    close_value = 'Close Value'
  )
```

Question 4

```
glimpse(dj_new)
```

```
## Rows: 1,492
## Columns: 2
## $ date        <chr> "05/29/1896", "06/30/1896", "07/31/1896", "08/31/1896",...
## $ close_value <dbl> 40.63, 36.15, 32.02, 31.97, 36.05, 39.53, 41.00, 40.45,...
```

```
#Effectve_date values are character and Close_Value values are numeric
```

Question 5

```
dj_new$month <- mdy(dj_new$date) %>%
  month()
```

Question 6

```
v <- diff(dj_new$month, lag = 1)
v[(v != 1) & (v!= -11)]
```

```
## [1] 5
```

```
dj_new$diff <- c(NA,v)
date <- dj_new[dj_new$diff != 1 & dj_new$diff != -11,]$date[2]
incident <- which(dj_new$date == date, arr.ind=TRUE)
dj_new[c(incident-1, incident, incident+1),]
```

```
## # A tibble: 3 x 4
##   date       close_value month  diff
##   <chr>            <dbl> <dbl> <dbl>
## 1 07/30/1914        71.4     7     1
## 2 12/31/1914        54.6    12     5
## 3 01/30/1915        57.2     1   -11
```

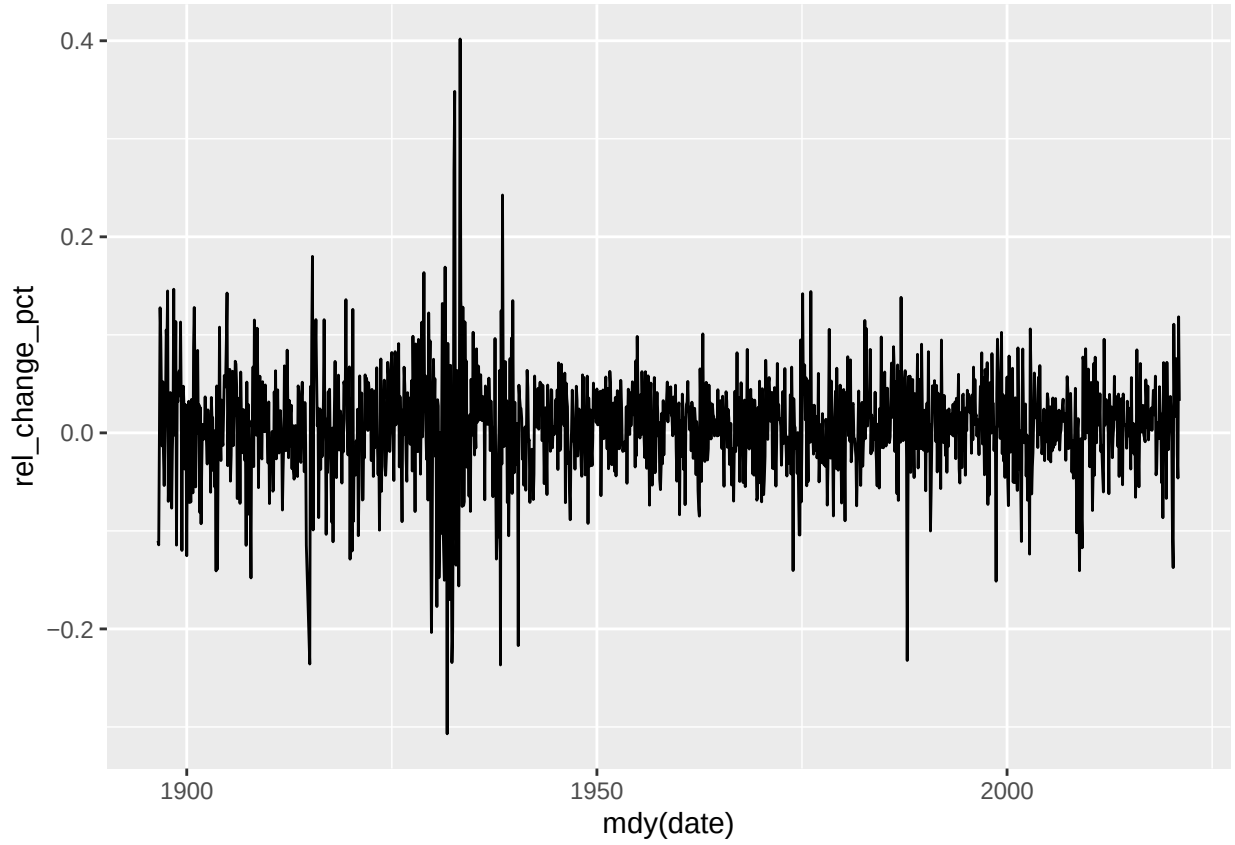
We find that there was a gap of 5 months between July 30th 1914 and December 31st 1914. Because of World War 1, all major European stock markets closed, which explains the gaps in the data of monthly performance history of Dow Jones from August to December of 1914.

Question 7

```
dj_new$rel_change_pct <- c(NA, diff(dj_new$close_value, lag = 1))
dj_new$rel_change_pct[2:nrow(dj_new)] <-
  dj_new$rel_change_pct[2:nrow(dj_new)]/dj_new$close_value[1:(nrow(dj_new)-1)]
```

Question 8

```
ggplot(dj_new[-1,], aes(mdy(date), rel_change_pct)) + geom_line()
```



Question 9

```
#dj_new[dj_new$rel_change_pct==max(dj_new$rel_change_pct),]
max(dj_new$rel_change_pct,na.rm = TRUE)
```

```
## [1] 0.4018051
```

Question 10

Question 11