

Intro to Data Science - Lab 6

IST687 Section M003

Professor Anderson

```
# Enter your name here: Hrishikesh Telang
```

```
###Select one of the below and add needed information
```

```
# 1. I did this homework by myself, with help from the book and the professor.
```

```
#Here are two lines of starter code:
```

```
library(ggplot2)
myPlot <- ggplot(economics, aes(x=date))
myPlot <- myPlot + geom_line(aes(y=psavert))
```

#1. Run these two lines of code. What happens? How do you actually “invoke” the plot (i.e., how do you get it to draw in the plot window)?

```
#The first line creates a ggplot function that just produces a visualization of
#'economics' with x axis as date
#With the myPlot already created, the second line overrides myPlot by adding
#psavert as the y axis
#We can invoke the plot by calling myPlot directly, like this:
myPlot
```



#2. Run `help("economics")` to find out the meaning of the `psavert` variable.

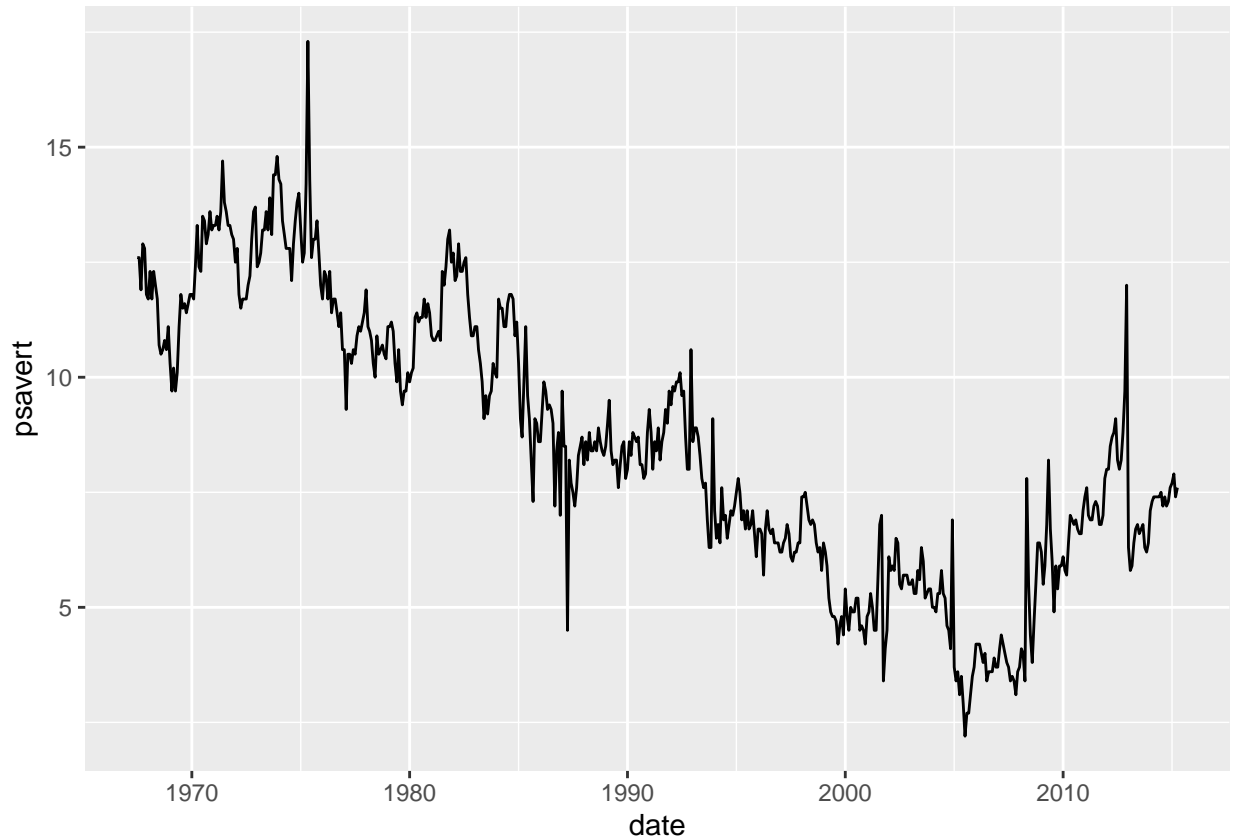
```
View(economics)
str(economics)
```

```
## spec_tbl_df [574 x 6] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ date      : Date[1:574], format: "1967-07-01" "1967-08-01" ...
## $ pce       : num [1:574] 507 510 516 512 517 ...
## $ pop       : num [1:574] 198712 198911 199113 199311 199498 ...
## $ psavert   : num [1:574] 12.6 12.6 11.9 12.9 12.8 11.8 11.7 12.3 11.7 12.3 ...
## $ uempmed   : num [1:574] 4.5 4.7 4.6 4.9 4.7 4.8 5.1 4.5 4.1 4.6 ...
## $ unemploy  : num [1:574] 2944 2945 2958 3143 3066 ...
```

```
help("economics")
```

#3. Examine the plot to estimate when the personal savings rate reached its maximum value. Also examine the plot to estimate when the personal savings rate reached its minimum value.

```
myPlot
```



*#From the plot, we can observe that the personal savings rate reached the peak
#somewhere close to 1975 and reached the lowest dip around 2005 to 2006*

#4. Use `which.max()` and `which.min()` to verify your guesses from problem 3.

```
#head(economics)
economics[which.max(economics$psavert),1]
```

```
## # A tibble: 1 x 1
##   date
##   <date>
## 1 1975-05-01
```

```
economics[which.min(economics$psavert),1]
```

```
## # A tibble: 1 x 1
##   date
##   <date>
## 1 2005-07-01
```

#5. Change the color of the plot line to green. Hint: Changing a line to a constant color happens in the specification of the geometry.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v tibble  3.1.4    v dplyr   1.0.7  
## v tidyr   1.1.3    v stringr 1.4.0  
## v readr   2.0.1    v forcats 0.5.1  
## v purrr   0.3.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

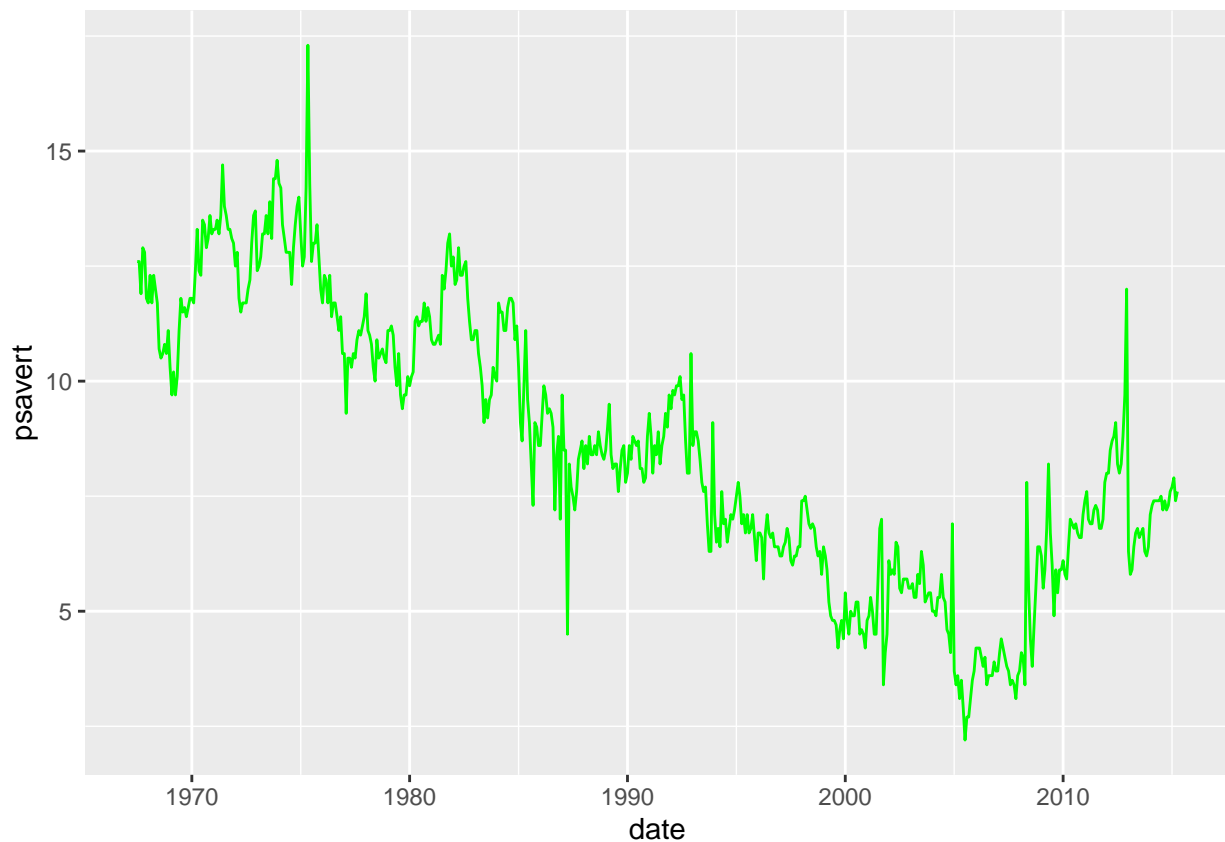
```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
myPlot <- ggplot(economics, aes(x=date))
```

```
myPlot <- myPlot + geom_line(aes(y=psavert), colour='green')
```

```
myPlot
```



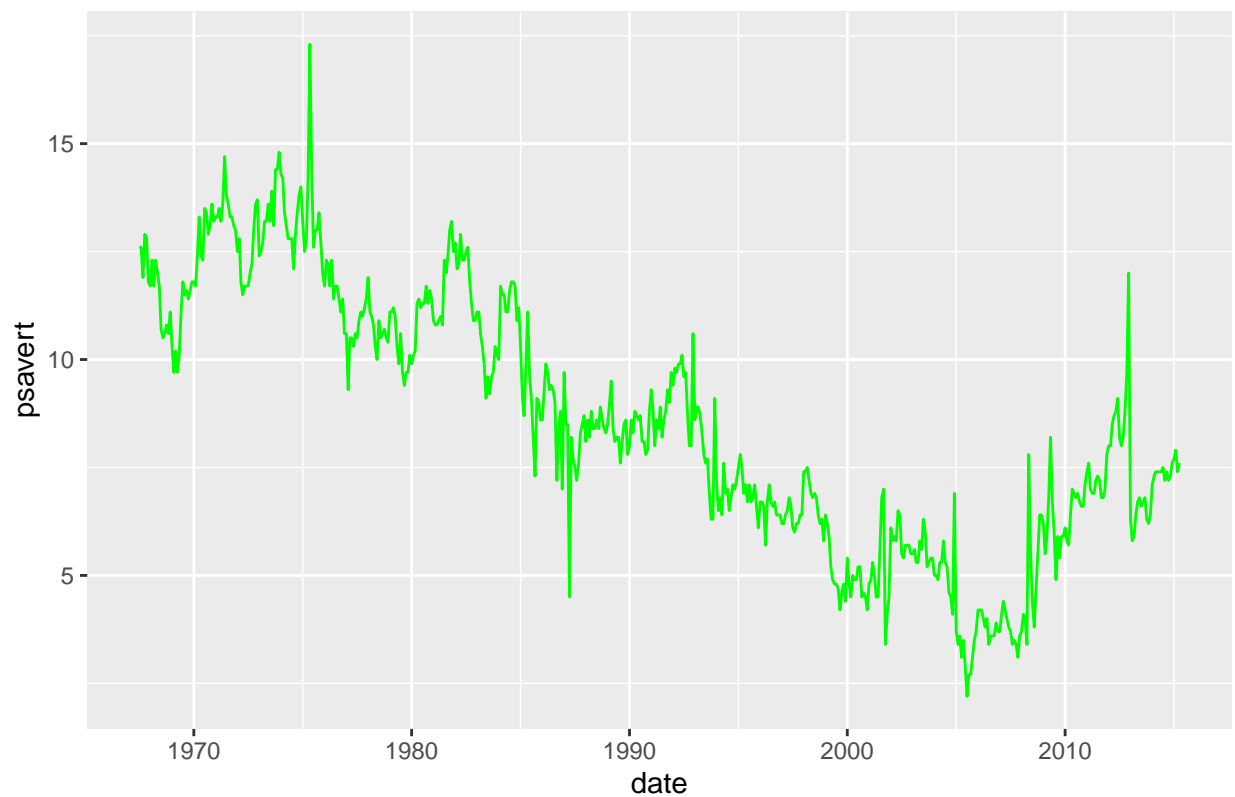
#6. Add a title to the plot with the ggtitle("Put title here") sub-command. The title "Personal Savings Rate: 1967-2014" would be a good choice.

```
myPlot <- ggplot(economics, aes(x=date))
```

```
myPlot <- myPlot + ggtitle("Personal Savings Rate: 1967-2014") + geom_line(aes(y=psavert), colour='green')
```

```
myPlot
```

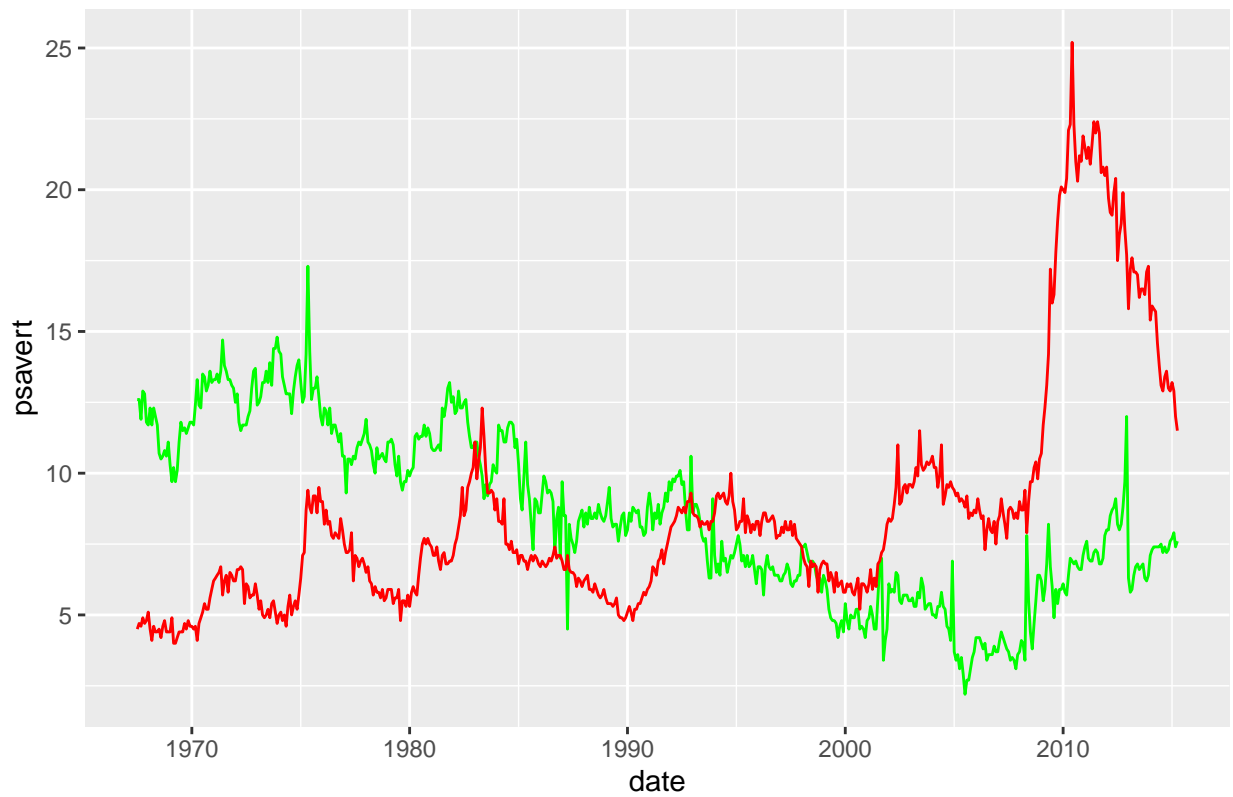
Personal Savings Rate: 1967–2014



#7. Add another data series to your plot to show the variable uempmed as a red line.

```
myPlot <- ggplot(economics, aes(x=date))  
myPlot <- myPlot + ggtitle("Personal Savings Rate: 1967-2014") + geom_line(aes(y=psavert), colour='green')  
myPlot
```

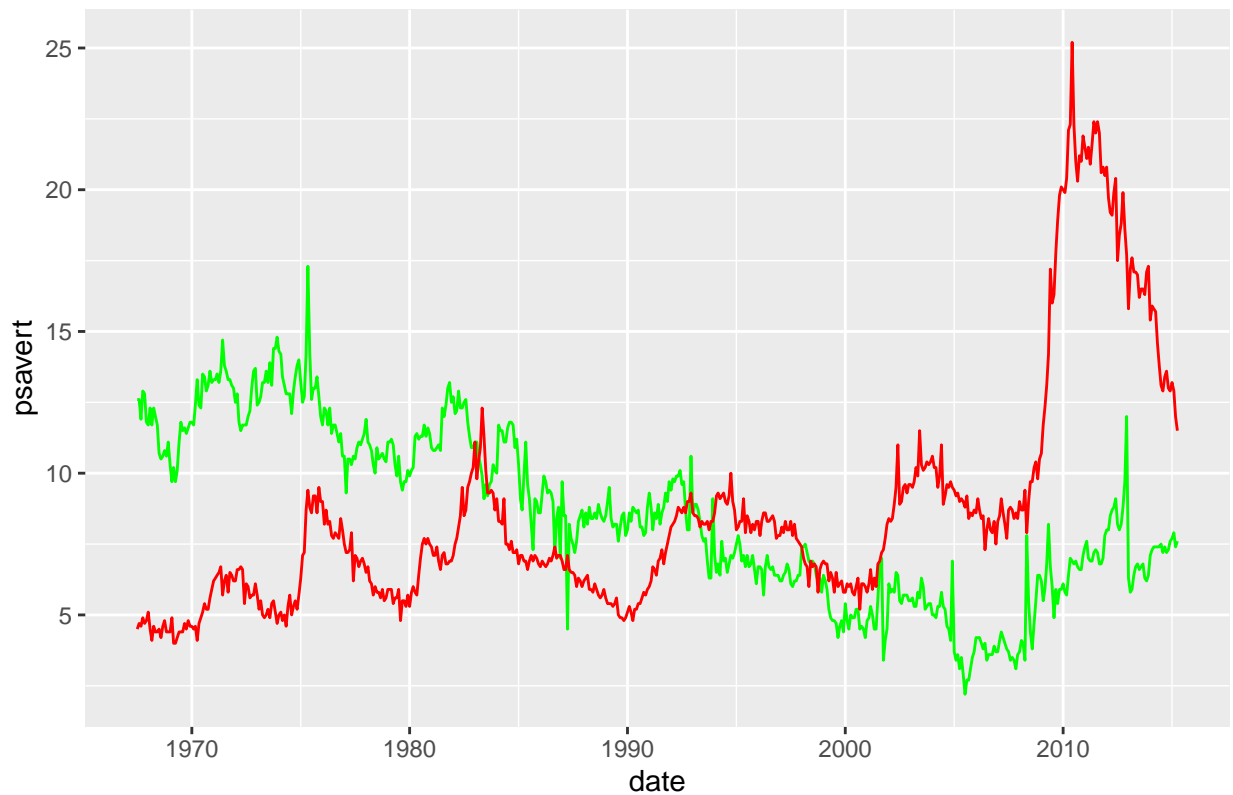
Personal Savings Rate: 1967–2014



#8. Change the title of the plot to mention both variables.

```
myPlot <- ggplot(economics, aes(x=date))  
myPlot <- myPlot + ggtitle("Personal Savings Rate and Median Duration of Unemployment: 1967-2014") + ge  
myPlot
```

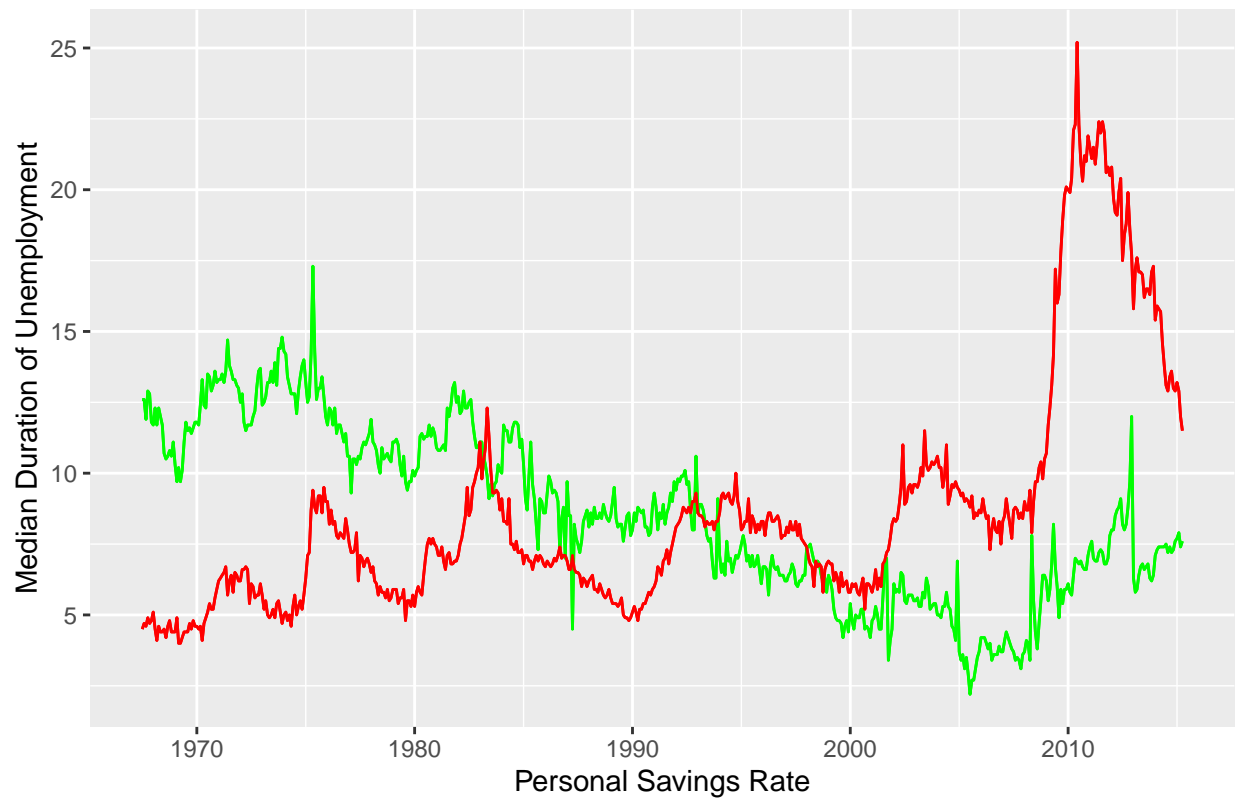
Personal Savings Rate and Median Duration of Unemployment: 1967–2014



#9. You can modify the axis labels in a ggplot with `ylab()` and `xlab()` subcommands. Change the axis labeling as needed to account for plotting both `psavert` and `uempmed` in the same window.

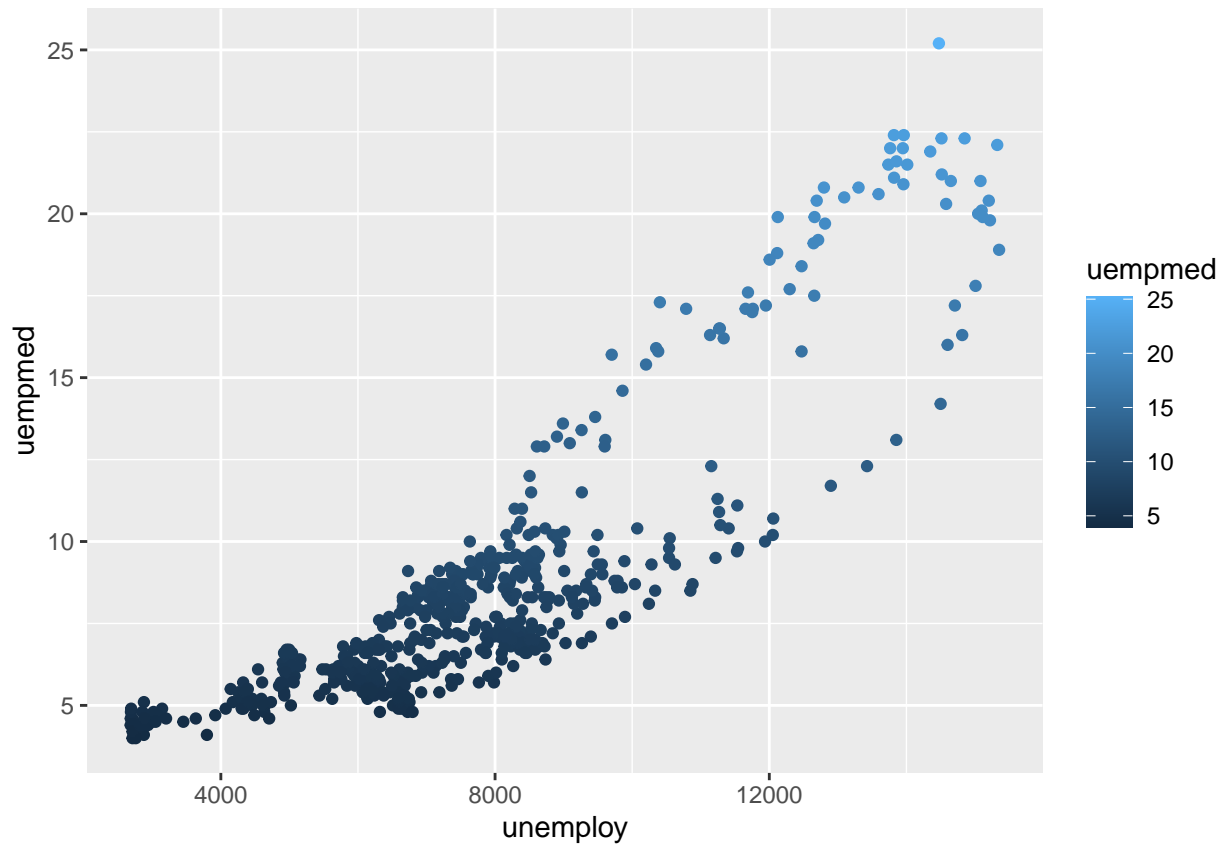
```
myPlot <- myPlot + ggtitle("Personal Savings Rate and Median Duration of Unemployment: 1967-2014") + ge  
myPlot
```

Personal Savings Rate and Median Duration of Unemployment: 1967–2014



#10. Create one last plot, creating a scatter plot, having the unemployment on the x-axis, psavert on the y-axis. Color each point based on the uempmed.

```
myPlot2 <- ggplot(economics, aes(x=unemploy)) + geom_point(aes(y=uempmed, colour=uempmed))  
myPlot2
```

#11. Interpret what you see in this last graph

*#It can be observed that most of the proportion of unemployment is highly concentrated
#between 2 to 10 weeks. However, there are a few outliers of unemployment cases
#beyond 20 weeks. It is not clear which industries have lesser weeks of
#unemployments.*

#12. Make sure your code has nice comments and an attribution statement!