

lab6

chaithra

07/10/2021

```
## Intro to Data Science - Lab 6

# IST687 Section M002
# Professor Anderson
# Enter your name here: Chaithra Koppram Cheluvaiah
# 1. I did this homework by myself, with help from the book and the professor.
```

```
library(ggplot2) # loading the ggplot2 package for data visualization

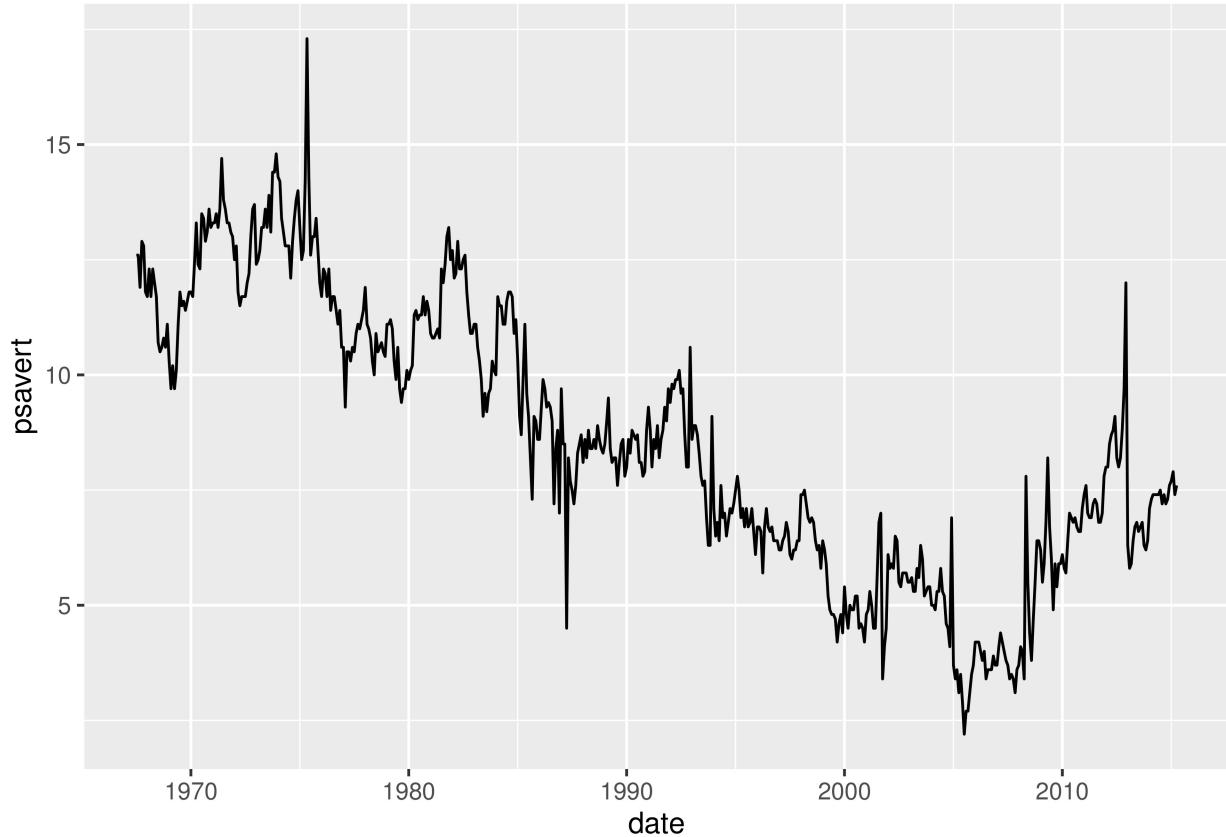
# plot with economics data and x-axis as date
myPlot <- ggplot(economics, aes(x=date))

# adding the geometry (time series chart)
# aesthetics to plot y-axis as psavert
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)?

```
# first line: creates empty plot with x-axis as date
# second line: adds the geometry as time series and y-axis as psavert

myPlot #to invoke the plot to draw
```



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

```
help("economics") # exploring data set
```

```
## starting httpd help server ... done
```

```
View(economics) # exploring data set
```

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.

```
# min around 2005
# max around 1975
```

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

```
# finds max row index
# fetches max row from data frame
economics[which.max(economics$psavert),]
```

```
## # A tibble: 1 x 6
##   date      pce    pop psavert uempmed unemploy
##   <date>    <dbl>  <dbl>    <dbl>    <dbl>    <dbl>
## 1 1975-05-01 1019. 215523     17.3     9.4     8433
```

```

# finds min row index
# fetches min row from data frame
economics[which.min(economics$psavert),]

## # A tibble: 1 x 6
##   date      pce    pop psavert uempmed unemploy
##   <date>    <dbl>  <dbl>    <dbl>    <dbl>
## 1 2005-07-01 8830. 296186     2.2     8.8     7406

summary(economics) # verifying min and max from summary as well

```

```

##       date           pce          pop      psavert
##   Min.   :1967-07-01   Min.   : 506.7   Min.   :198712   Min.   : 2.200
##  1st Qu.:1979-06-08   1st Qu.: 1578.3   1st Qu.:224896   1st Qu.: 6.400
##  Median :1991-05-16   Median : 3936.8   Median :253060   Median : 8.400
##  Mean   :1991-05-17   Mean   : 4820.1   Mean   :257160   Mean   : 8.567
##  3rd Qu.:2003-04-23   3rd Qu.: 7626.3   3rd Qu.:290291   3rd Qu.:11.100
##  Max.   :2015-04-01   Max.   :12193.8   Max.   :320402   Max.   :17.300
##       uempmed        unemploy
##   Min.   : 4.000   Min.   : 2685
##  1st Qu.: 6.000   1st Qu.: 6284
##  Median : 7.500   Median : 7494
##  Mean   : 8.609   Mean   : 7771
##  3rd Qu.: 9.100   3rd Qu.: 8686
##  Max.   :25.200   Max.   :15352

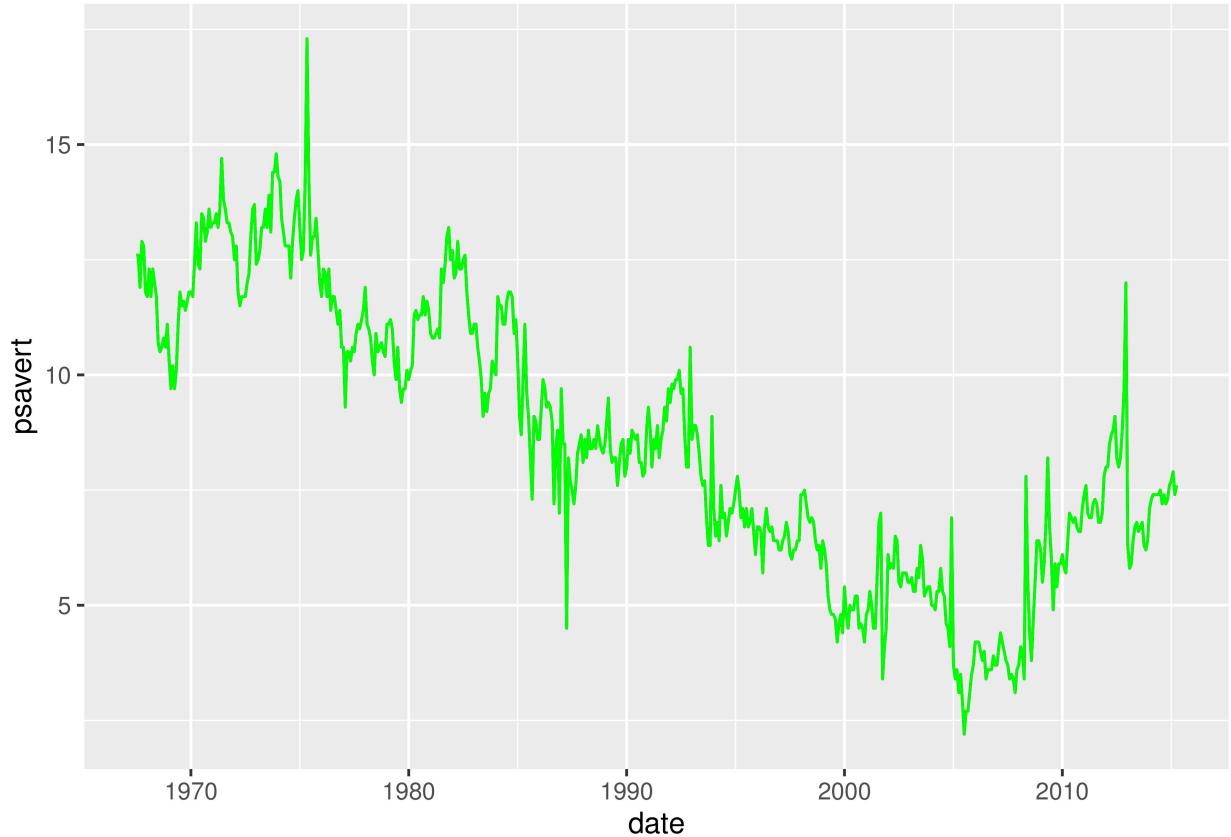
```

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.

```

# col attribute to change the color of the trend line
myPlot <- myPlot + geom_line(aes(y=psavert), col="green")
myPlot

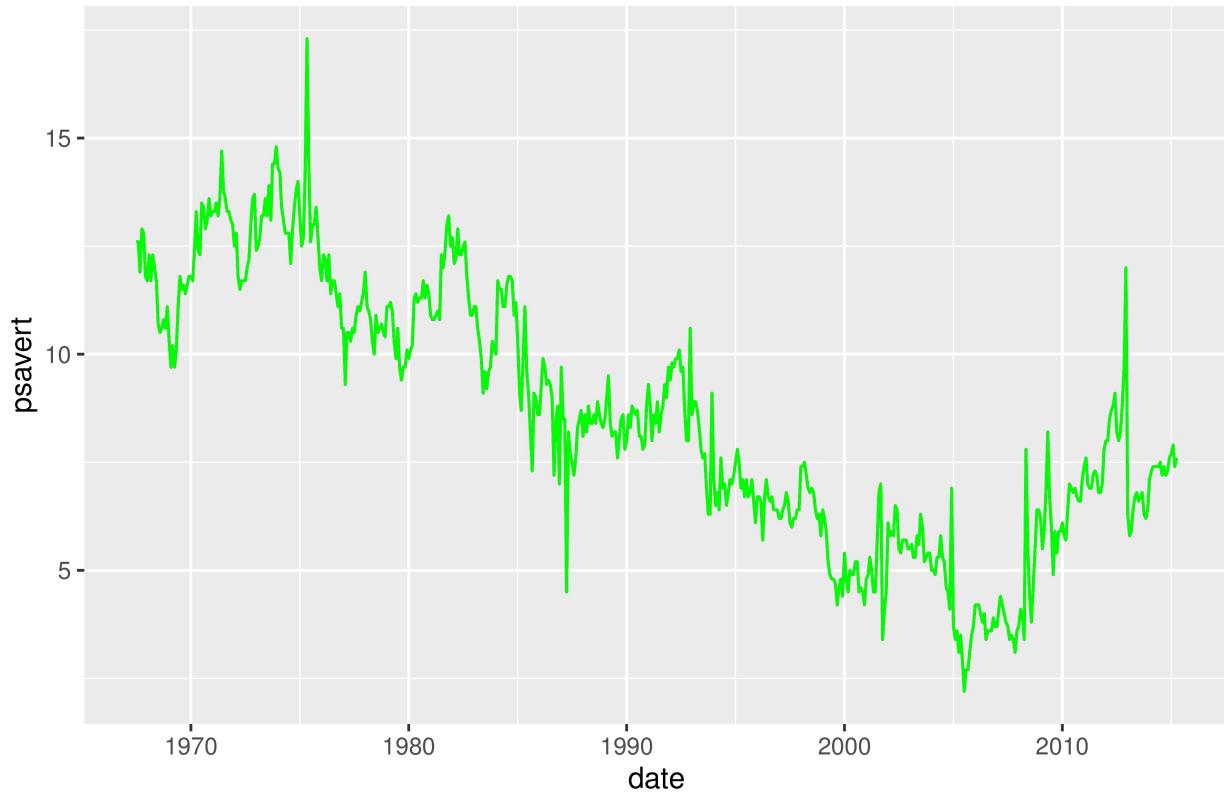
```



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.

```
# adding title to the graph
myPlot <- myPlot + ggtitle("Personal Savings Rate: 1967-2014")
myPlot
```

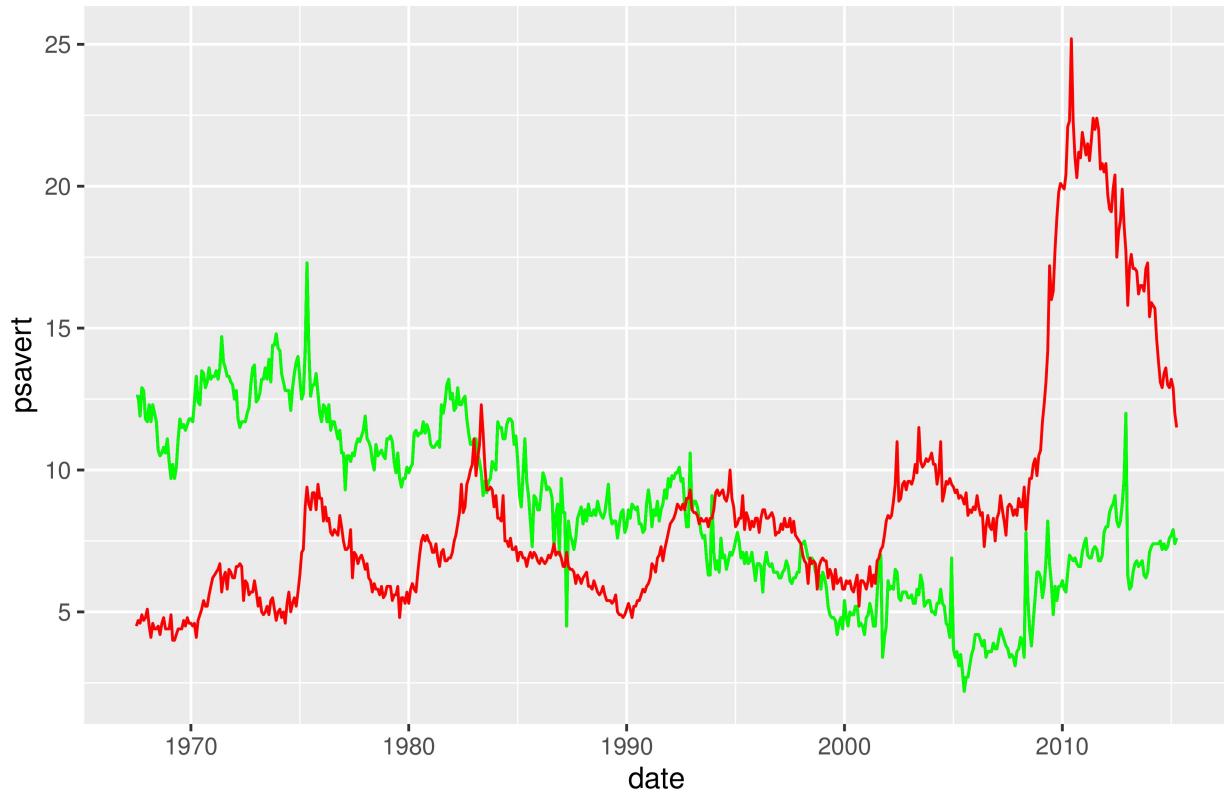
Personal Savings Rate: 1967–2014



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

```
# adding trend line for median unemployed weeks column
myPlot <- myPlot + geom_line(aes(y=uempmed), col="red")
myPlot
```

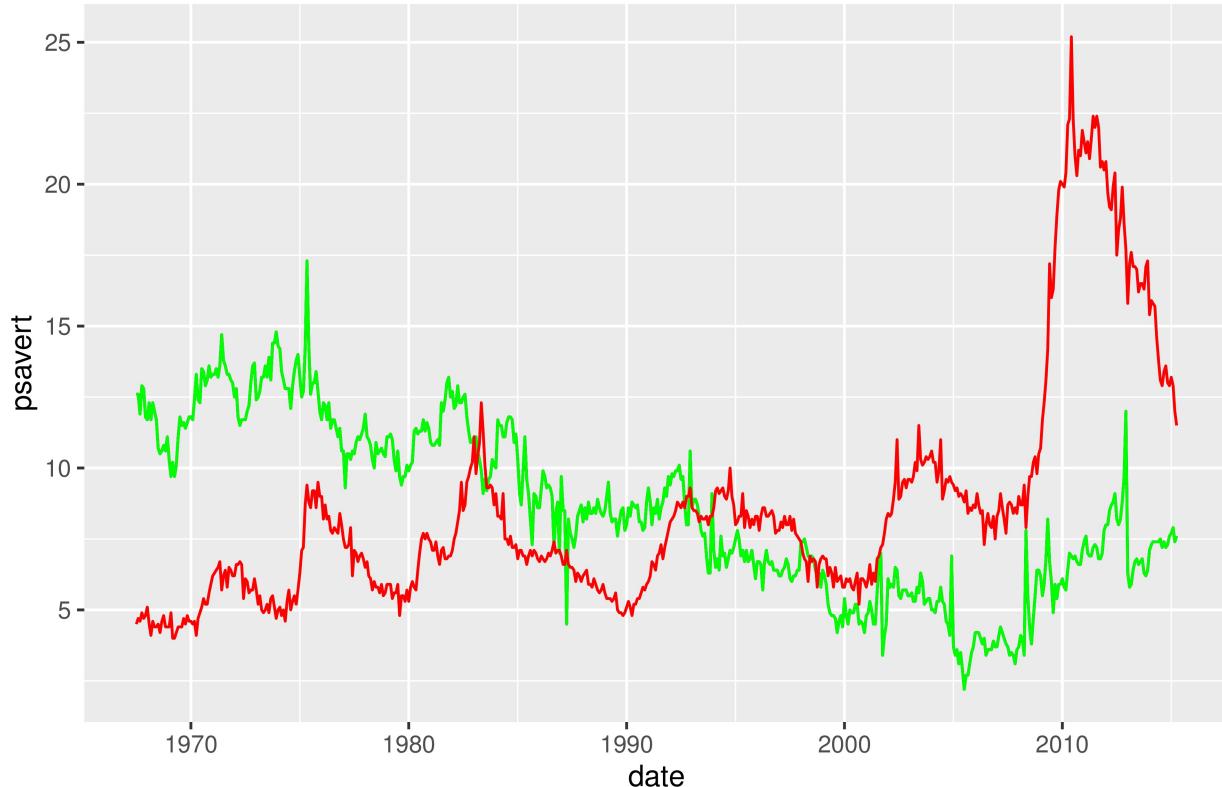
Personal Savings Rate: 1967–2014



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

```
# updating the title
myPlot <- myPlot + ggtitle("Personal Savings and Median Unemployment Rate: 1967-2014")
myPlot
```

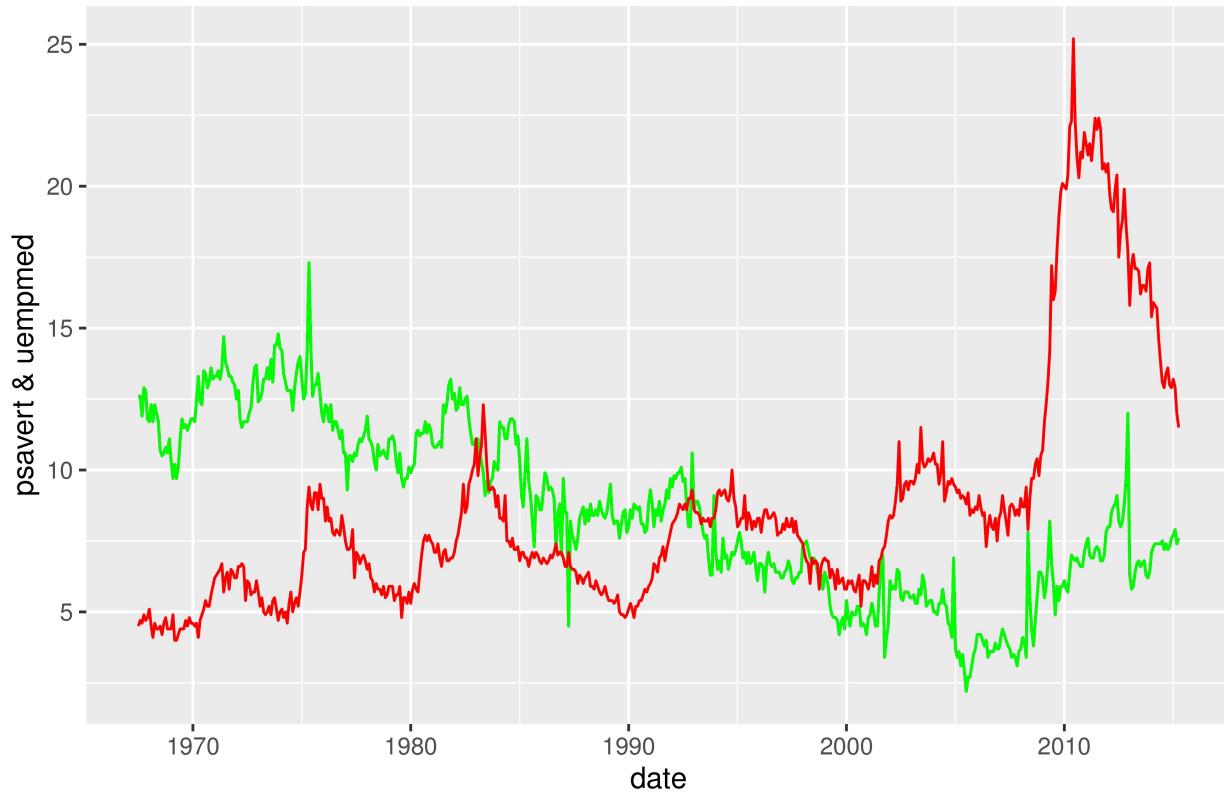
Personal Savings and Median Unemployment Rate: 1967–2014



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.

```
# changing y-axis title
myPlot <- myPlot + ylab("psavert & uempmed")
myPlot
```

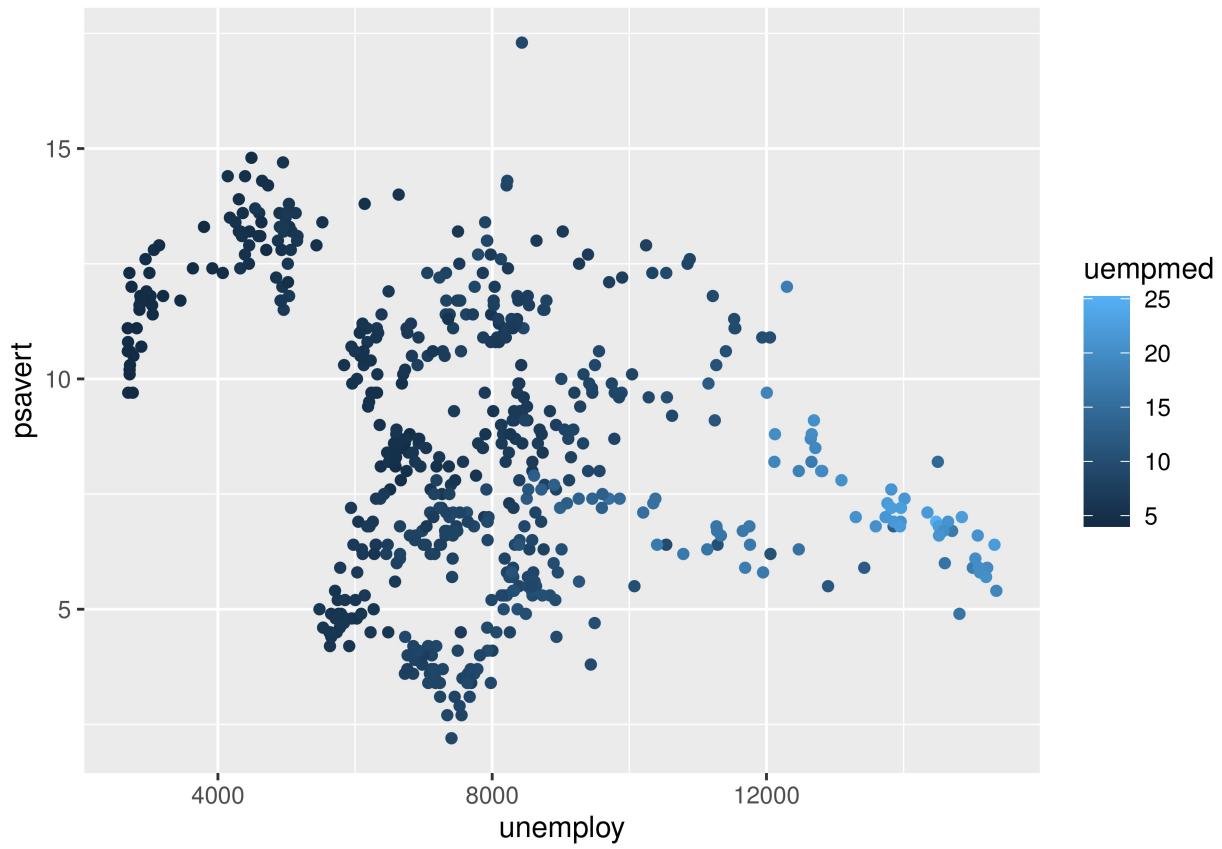
Personal Savings and Median Unemployment Rate: 1967–2014



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

```
# creating plot with economics data
# aesthetics with x-axis as unemployment level and y-axis as psavert
scatter <- ggplot(economics, aes(x=unemploy, y=psavert))

# scatter plot with intensity of the color defined by uempmed variable
scatter <- scatter + geom_point(aes(color=uempmed))
scatter # invoking the plot
```



11. Interpret what you see in this last graph

```
# above scatter plot represents plots three numeric variables
# x-axis: number of people unemployed
# y-axis: personal savings rate of unemployed people
# third numeric variable uempmed (median weeks unemployed) is visualized by
# intensity of the color

# Inference from the graph:
# people who are unemployed for more number of weeks have low personal savings
```

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