Multi Variable Time Series Plot - ggplot - practice

Monica Buczynski

8/20/2021

```
colnames(unem)
```

```
##  [1] "Year" "Jan"  "Feb"  "Mar"  "Apr"  "May"  "Jun"  "Jul"  "Aug"  "Sep"
## [11] "Oct"  "Nov"  "Dec"
```

```
colnames(psvrt)
```

```
## [1] "date"    "psavert"
```

```
summary(unem)
```

```
##       Year           Jan              Feb             Mar
##  Min.   :1948   Min.   : 2.900   Min.   : 2.6   Min.   : 2.600
##  1st Qu.:1966   1st Qu.: 4.600   1st Qu.: 4.6   1st Qu.: 4.400
##  Median :1984   Median : 5.600   Median : 5.5   Median : 5.500
##  Mean   :1984   Mean   : 5.705   Mean   : 5.7   Mean   : 5.719
##  3rd Qu.:2002   3rd Qu.: 6.600   3rd Qu.: 6.6   3rd Qu.: 6.700
##  Max.   :2020   Max.   :10.400   Max.   :10.4   Max.   :10.300
##
##       Apr              May              Jun              Jul
##  Min.   : 2.700   Min.   : 2.500   Min.   : 2.500   Min.   :2.600
##  1st Qu.: 4.600   1st Qu.: 4.400   1st Qu.: 4.600   1st Qu.:4.575
##  Median : 5.600   Median : 5.600   Median : 5.500   Median :5.500
##  Mean   : 5.859   Mean   : 5.836   Mean   : 5.822   Mean   :5.740
##  3rd Qu.: 7.000   3rd Qu.: 7.100   3rd Qu.: 7.000   3rd Qu.:6.900
##  Max.   :14.700   Max.   :13.300   Max.   :11.100   Max.   :9.800
##                                                     NA's   :1
##       Aug              Sep              Oct              Nov
##  Min.   :2.700   Min.   : 2.900   Min.   : 3.000   Min.   : 2.800
##  1st Qu.:4.500   1st Qu.: 4.475   1st Qu.: 4.475   1st Qu.: 4.475
##  Median :5.600   Median : 5.500   Median : 5.550   Median : 5.650
##  Mean   :5.733   Mean   : 5.718   Mean   : 5.735   Mean   : 5.726
##  3rd Qu.:6.825   3rd Qu.: 6.725   3rd Qu.: 6.800   3rd Qu.: 6.650
##  Max.   :9.800   Max.   :10.100   Max.   :10.400   Max.   :10.800
##  NA's   :1       NA's   :1        NA's   :1        NA's   :1
##       Dec
##  Min.   : 2.700
##  1st Qu.: 4.475
##  Median : 5.500
##  Mean   : 5.733
##  3rd Qu.: 6.600
##  Max.   :10.800
##  NA's   :1
```

```r
summary(psvrt)
```

```
##       date         psavert     
##  Min.   :1959   Min.   : 3.175  
##  1st Qu.:1974   1st Qu.: 6.758  
##  Median :1989   Median : 8.800  
##  Mean   :1989   Mean   : 8.822  
##  3rd Qu.:2004   3rd Qu.:11.200  
##  Max.   :2019   Max.   :13.475
```

```r
# sum of NA's in data frame

sum(is.na(unem))
```

```
## [1] 6
```

```r
sum(is.na(psvrt))
```

```
## [1] 0
```

```r
# extracting positions of NA's in unem

which(is.na(unem), arr.ind = TRUE)
```

```
##      row col
## [1,]  73   8
## [2,]  73   9
## [3,]  73  10
## [4,]  73  11
## [5,]  73  12
## [6,]  73  13
```

```r
# print row 73

unem[73,]
```

```
## # A tibble: 1 x 13
##    Year   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2020   3.6   3.5   4.4  14.7  13.3  11.1    NA    NA    NA    NA    NA    NA
```

```r
# remove rows with NA's

unem2 <- na.omit(unem)

# check NA's in unem2 df

sum(is.na(unem2))
```

```
## [1] 0
```

```r
# dimensions

dim(unem2)
```

```
## [1] 72 13
```

```r
dim(psvrt)
```

```
## [1] 61  2
```

```
# psvrt
# unem2
```

```
# uppercase to lowercase column names

colnames(unem2) <- tolower(colnames(unem2))

# check
unem2
```

```
## # A tibble: 72 x 13
##     year   jan   feb   mar   apr   may   jun   jul   aug   sep   oct   nov   dec
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  1948   3.4   3.8   4     3.9   3.5   3.6   3.6   3.9   3.8   3.7   3.8   4
##  2  1949   4.3   4.7   5     5.3   6.1   6.2   6.7   6.8   6.6   7.9   6.4   6.6
##  3  1950   6.5   6.4   6.3   5.8   5.5   5.4   5     4.5   4.4   4.2   4.2   4.3
##  4  1951   3.7   3.4   3.4   3.1   3     3.2   3.1   3.1   3.3   3.5   3.5   3.1
##  5  1952   3.2   3.1   2.9   2.9   3     3     3.2   3.4   3.1   3     2.8   2.7
##  6  1953   2.9   2.6   2.6   2.7   2.5   2.5   2.6   2.7   2.9   3.1   3.5   4.5
##  7  1954   4.9   5.2   5.7   5.9   5.9   5.6   5.8   6     6.1   5.7   5.3   5
##  8  1955   4.9   4.7   4.6   4.7   4.3   4.2   4     4.2   4.1   4.3   4.2   4.2
##  9  1956   4     3.9   4.2   4     4.3   4.3   4.4   4.1   3.9   3.9   4.3   4.2
## 10  1957   4.2   3.9   3.7   3.9   4.1   4.3   4.2   4.1   4.4   4.5   5.1   5.2
## # ... with 62 more rows
```

```
colnames(unem2)
```

```
##  [1] "year" "jan"  "feb"  "mar"  "apr"  "may"  "jun"  "jul"  "aug"  "sep"
## [11] "oct"  "nov"  "dec"
```

```
# unem3 <- unem2 %>% pivot_longer(cols = c(2:13), names_to = "month", values_to = "value")

# unem2 %>% mutate(mean_unem = mean(2:13))

unem2$mean_unem = rowMeans(unem2[,c(-1)])

colnames(unem2)
```

```
##  [1] "year"      "jan"       "feb"       "mar"       "apr"       "may"
##  [7] "jun"       "jul"       "aug"       "sep"       "oct"       "nov"
## [13] "dec"       "mean_unem"
```

```
unem3 <- unem2 %>%
  select("year", "mean_unem")
```

```
# join the two tables together

df1 <-  psvrt %>%
  rename("year" = "date") %>%
  inner_join(unem3, by = "year") %>%
  rename("mean_psavrt" = "psavert") %>%
  pivot_longer(cols = c(2:3), values_to = "value", names_to = "variable")

df1
```

```
## # A tibble: 122 x 3
##     year variable    value
```

```
##    <dbl> <chr>       <dbl>
##  1  1959 mean_psavrt 10.3
##  2  1959 mean_unem    5.45
##  3  1960 mean_psavrt 10.1
##  4  1960 mean_unem    5.54
##  5  1961 mean_psavrt 11.3
##  6  1961 mean_unem    6.69
##  7  1962 mean_psavrt 11.2
##  8  1962 mean_unem    5.57
##  9  1963 mean_psavrt 10.7
## 10  1963 mean_unem    5.64
## # ... with 112 more rows
```

```r
# ggplot

ggplot(df1, aes(y = value, x = year)) +
  geom_line(aes(color = variable)) +
  scale_color_brewer(palette = "Dark2",
                     name = "Rates",
                     labels = c("Personal Saving", "Unemployment"))+
  labs(x = "Year",
       y = "Mean Annual Value",
       title = "Relationship between U.S. Annual Personal Savings Rate\n and Unemployment Rate",
       subtitle = "1959-2019",
       caption = "Data Sources: U.S. Bureau of Economic Analysis\n and U.S. Bureau of Labor Statistics")
  scale_y_continuous(breaks = seq(0,12,2))
```