Homework2

Jun Rao

8/15/2020

Homework 2: data visualization using ggplot2

These are exercises from

http://members.cbio.mines-paristech.fr/~thocking/animint2-manual/Ch02-ggplot2.html

(the variable name viz.aligned below refers to ggplot code on that web page).

```
#if we didn't in stll remotes package, we install it
if(FALSE){
    install.packages("remotes")
}

#if we didn't in stll animint2 package, we install it
if(FALSE){
    remotes::install_github("tdhock/animint2")
}

library(animint2)
```

1.In viz.aligned we showed a ggplot with a scatterplot panel on the left and a time series panel on the right. Make another version of the data visualization with the time series panel on the left and the scatterplot panel on the right. Hint: the order of panels in ggplots is controlled by the factor levels of the facet variable. For example facet_grid(. ~ x) means use the different values of the x variable in different facets, where x = factor(values, levels).

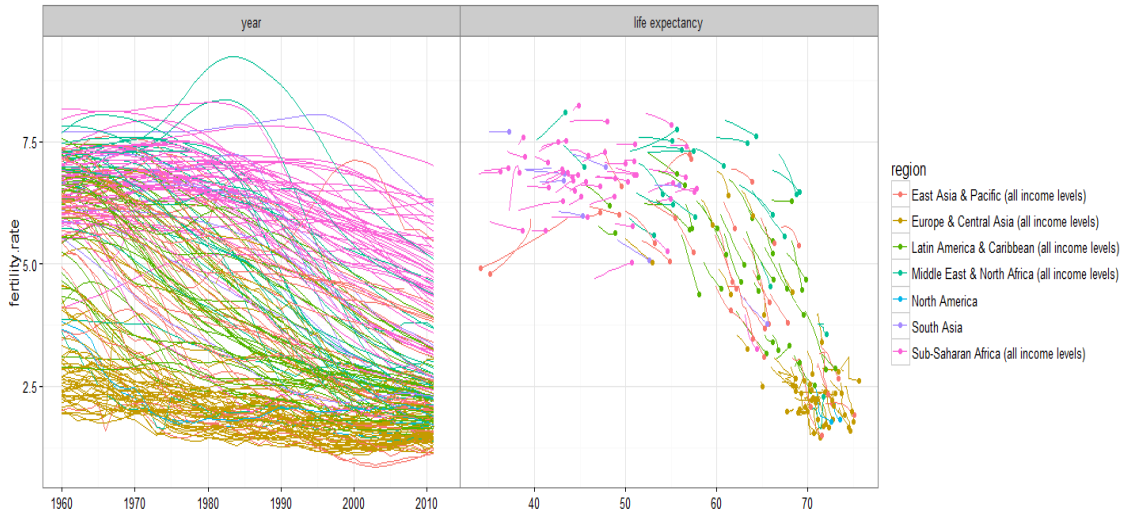
```
add.x.var <- function(df, x.var){
  data.frame(df, x.var=factor(x.var, c("year", "life expectancy")))
}


viz.aligned <- function(df1,df2,df){
  viz.aligned <- ggplot()+
      geom_path(aes(x=life.expectancy,y=fertility.rate,color=region,
                    group=country),
                data=add.x.var(df1, "life expectancy"))+
    geom_point(aes(x=life.expectancy,y=fertility.rate,color=region),
               data=add.x.var(df2, "life expectancy"))+
```

```
    geom_line(aes(x=year,y=fertility.rate, color=region, group=country),
              data=add.x.var(df, "year"))+
    xlab("")+
    facet_grid(. ~ x.var, scales="free")+
    theme_bw()+
    theme(panel.margin=grid::unit(0, "lines"))


  return(viz.aligned)
}


viz.aligned(subset(WorldBank, 1970 <= year & year <= 1975),subset(WorldBank,
year==1975),WorldBank)
```



2.In viz. aligned the scatterplot displays fertility rate and life expectancy, but the time series displays only fertility rate. Make another version of the data visualization that shows both time series. There should be only one ggplot that looks something like this. Hint: add another geom_point and use both horizontal and vertical panels in facet_grid(row_var ~ col_var).

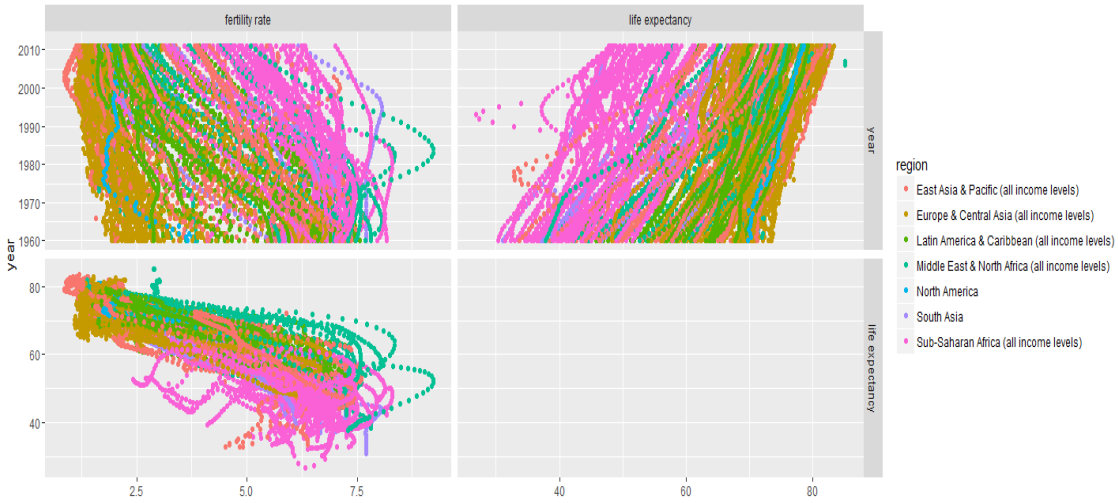
```
add.vars <- function(df,x.var,y.var){
  data.frame(df,x.var=factor(x.var, c("fertility rate", "life expectancy")),
y.var=factor(y.var, c("year", "life expectancy")))
}

viz.aligned2 <- function(df){
  result <- ggplot()+
```

```
    geom_point(aes(x=life.expectancy,y=year, color=region),
               data=add.vars(df, "life expectancy","year"))+
    geom_point(aes(x=fertility.rate,y=year, color=region),
              data=add.vars(df, "fertility rate","year"))+
     geom_point(aes(x=fertility.rate,y=life.expectancy, color=region),
              data=add.vars(df, "fertility rate","life expectancy"))+
    xlab("")+
    facet_grid(y.var~ x.var, scales="free")

  return(result)
}


viz.aligned2(WorldBank)
```



3.Create a multi-panel data visualization (scatterplot of data for one year in each panel) that shows each year of the WorldBank data set in a separate panel of one ggplot. Hint: use facet_wrap.

```
viz.panels <- function(df){
  result <-
ggplot(df,aes(life.expectancy,fertility.rate,color=region,group=country)) +
      geom_point() +
      facet_wrap(~year)


  return(result)
}


viz.panels(WorldBank)
```


1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
fertility.rate
7.5
5.0
2.5
40
60
80
life.expectancy
region
East Asia & Pacific (all income levels)
Europe & Central Asia (all income levels)
Latin America & Caribbean (all income levels)
Middle East & North Africa (all income levels)
North America
South Asia
Sub-Saharan Africa (all income levels)