Assignment 4

Min Shi

```r
## EPPS 6356 Assignment 4.R
## Min Shi
## Oct. 11th, 2021

# clear environment, set working directory and read the data
rm(list=ls())
setwd("/Users/min/Desktop/2021 Fall Semester/EPPS 6356 Data Visualization/Assignment 4/Results")

# load the libraries
library(aod)
library(ggplot2)
library(hrbrthemes) # for Chart 1 -- style of Variable Width Column Chart
library(dplyr) # for group_by function
library(reshape2) # for Chart 2 -- Table with Embedded Charts
library(viridis) # for Chart 3 -- Small multiple bar charts
library(hrbrthemes) # for Chart 3 -- Small multiple bar charts
library(qdap) # for Chart 4 -- character strip

# load the datasets
data1 <- read.csv("/Users/min/Desktop/2021 Fall Semester/EPPS 6356 Data Visualization/Assignment 4/population_by_
country_2020.csv")
data2 <- read.csv("/Users/min/Desktop/2021 Fall Semester/EPPS 6356 Data Visualization/Assignment 4/world-happines
s-report-2021.csv")
ls(data1)
```

```
##  [1] "Country..or.dependency." "Density..P.Km.."
##  [3] "Fert..Rate"              "Land.Area..Km.."
##  [5] "Med..Age"                "Migrants..net."
##  [7] "Net.Change"              "Population..2020."
##  [9] "Urban.Pop.."             "World.Share"
## [11] "Yearly.Change"
```

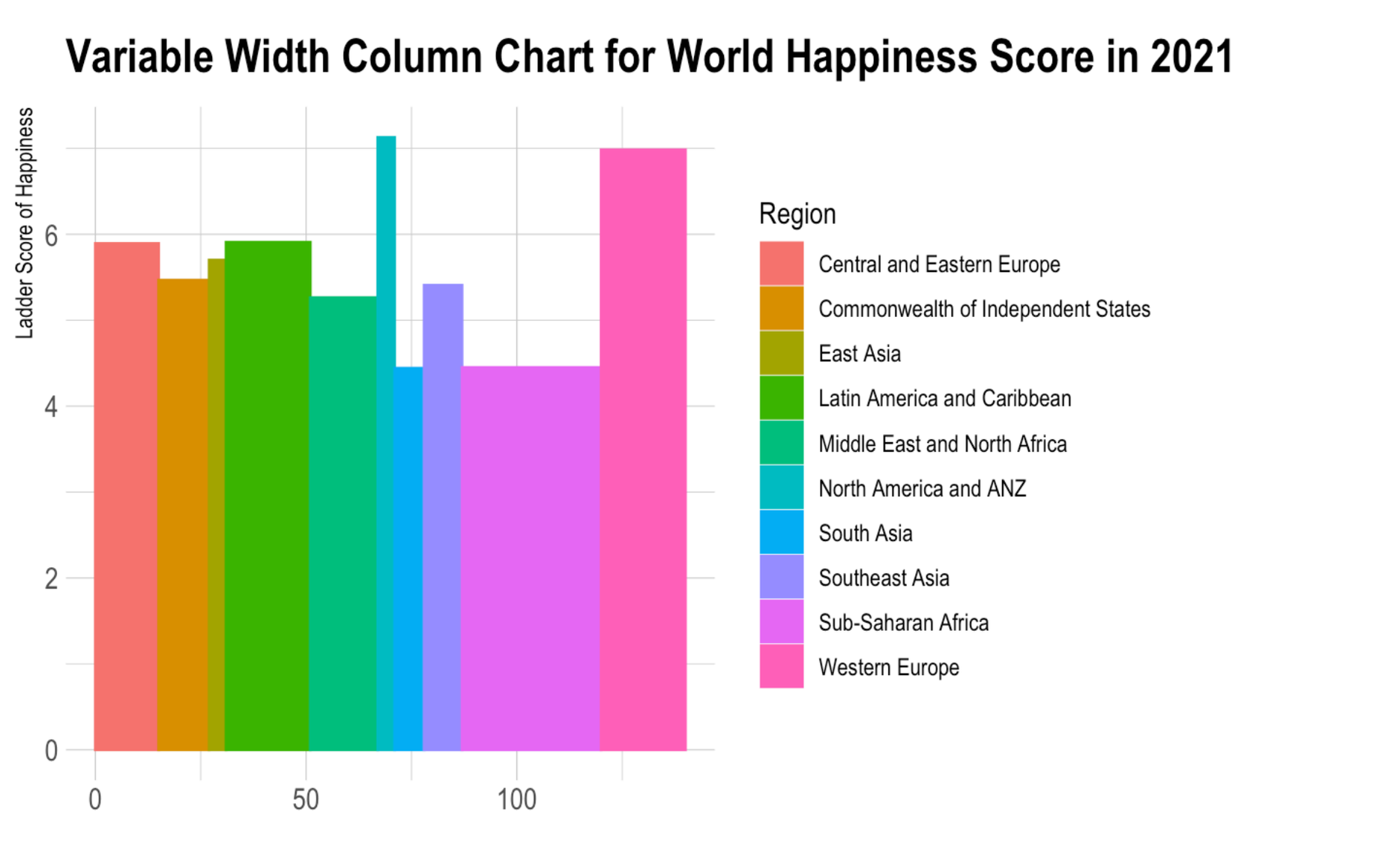
```r
ls(data2)
```

```
##  [1] "Country.name"
##  [2] "Dystopia...residual"
##  [3] "Explained.by..Freedom.to.make.life.choices"
##  [4] "Explained.by..Generosity"
##  [5] "Explained.by..Healthy.life.expectancy"
##  [6] "Explained.by..Log.GDP.per.capita"
##  [7] "Explained.by..Perceptions.of.corruption"
##  [8] "Explained.by..Social.support"
##  [9] "Freedom.to.make.life.choices"
## [10] "Generosity"
## [11] "Healthy.life.expectancy"
## [12] "Ladder.score"
## [13] "Ladder.score.in.Dystopia"
## [14] "Logged.GDP.per.capita"
## [15] "lowerwhisker"
## [16] "Perceptions.of.corruption"
## [17] "Regional.indicator"
## [18] "Social.support"
## [19] "Standard.error.of.ladder.score"
## [20] "upperwhisker"
```

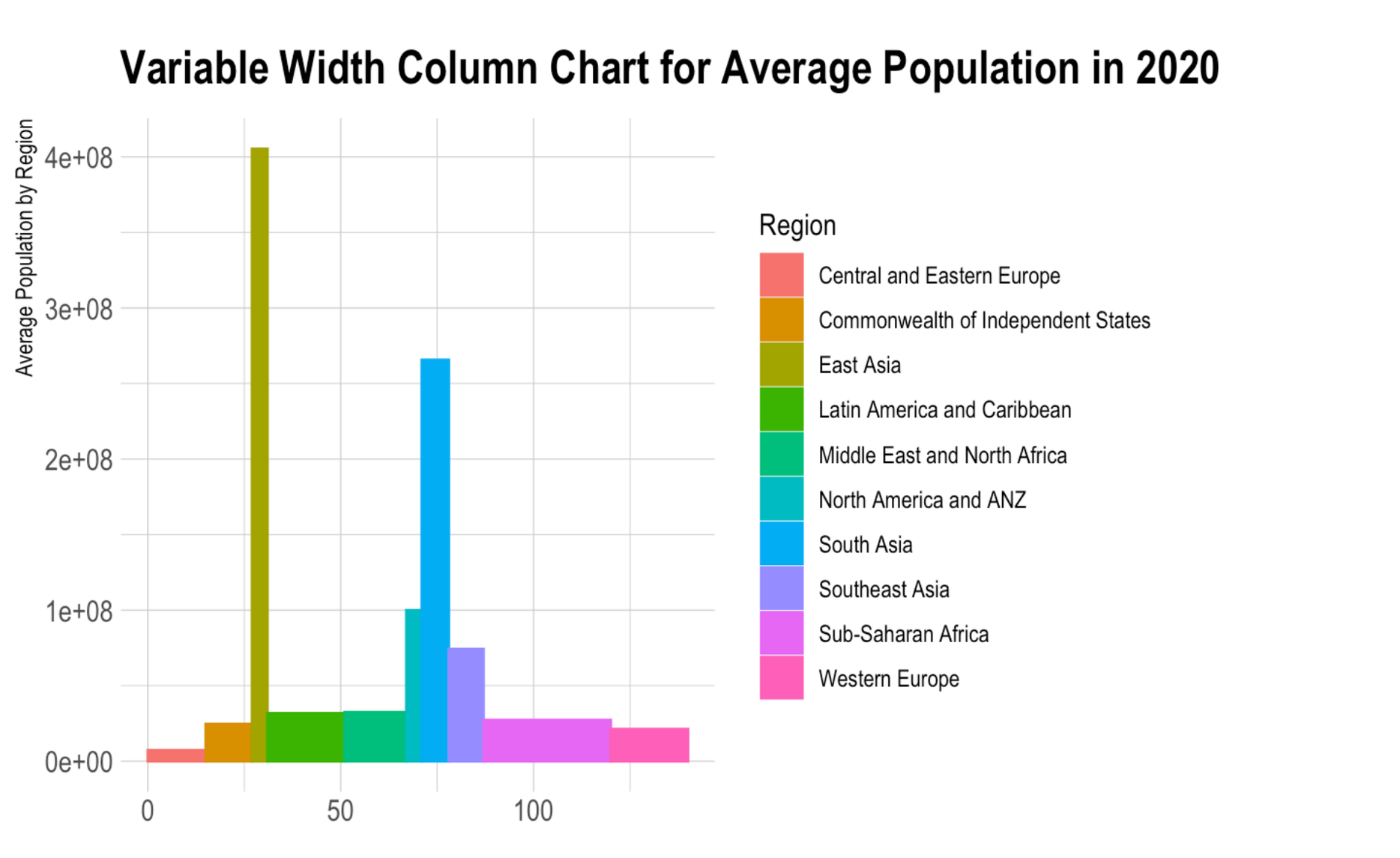
```r
data1 <- rename(data1, Country = Country..or.dependency.)
data2 <- rename(data2, Country = Country.name)
mydata <- merge(data1, data2, by="Country")


#####################################
# Chart 1 Variable Width Column Chart
#####################################
# Sample 1
mydata <- rename(mydata, Region = Regional.indicator)
ls(mydata)
```

```
##  [1] "Country"
##  [2] "Density..P.Km.."
##  [3] "Dystopia...residual"
##  [4] "Explained.by..Freedom.to.make.life.choices"
##  [5] "Explained.by..Generosity"
##  [6] "Explained.by..Healthy.life.expectancy"
##  [7] "Explained.by..Log.GDP.per.capita"
##  [8] "Explained.by..Perceptions.of.corruption"
##  [9] "Explained.by..Social.support"
## [10] "Fert..Rate"
## [11] "Freedom.to.make.life.choices"
## [12] "Generosity"
## [13] "Healthy.life.expectancy"
## [14] "Ladder.score"
## [15] "Ladder.score.in.Dystopia"
## [16] "Land.Area..Km.."
## [17] "Logged.GDP.per.capita"
## [18] "lowerwhisker"
## [19] "Med..Age"
## [20] "Migrants..net."
## [21] "Net.Change"
## [22] "Perceptions.of.corruption"
## [23] "Population..2020."
## [24] "Region"
## [25] "Social.support"
## [26] "Standard.error.of.ladder.score"
## [27] "upperwhisker"
## [28] "Urban.Pop.."
## [29] "World.Share"
## [30] "Yearly.Change"
```

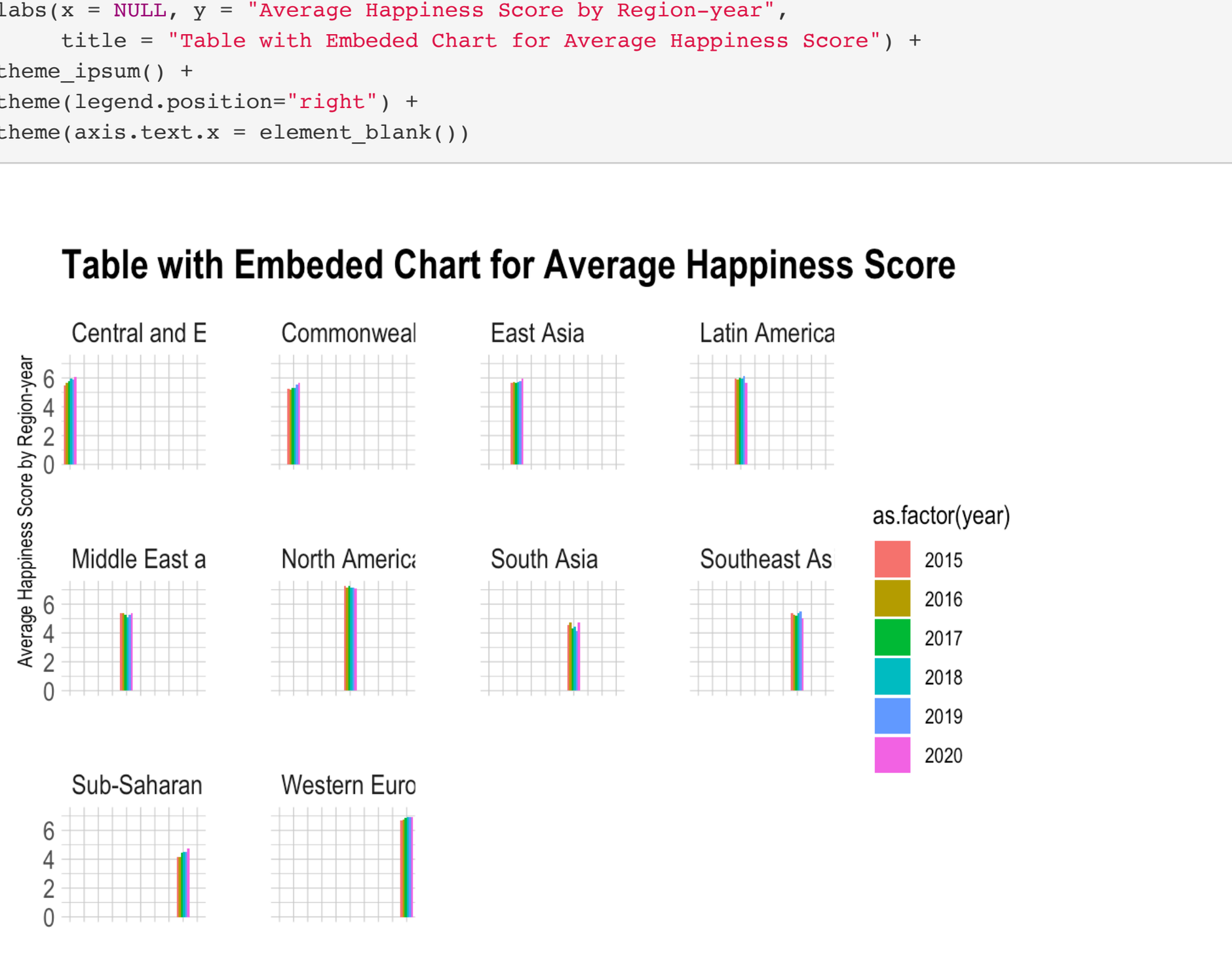
```r
mydata1_1 <- mydata %>% group_by(Region)
Ladder_by_region <- mydata1_1 %>% summarise(
  Ladder_score = mean(Ladder.score)
)

Number_of_states <- count(mydata1_1, 'Region')

Chart1_data <- merge(Ladder_by_region, Number_of_states, by="Region")
Chart1_data <- rename(Chart1_data, Number_of_states = n)

# Calculate the future positions on the x axis of each bar (left border, central position, right border)
Chart1_data$w <- cumsum(Chart1_data$Number_of_states)
Chart1_data$wm <- Chart1_data$w - Chart1_data$Number_of_states
Chart1_data$wt <- with(Chart1_data, wm + (w-wm)/2)

# Plot
ggplot(Chart1_data, aes(ymin = 0)) +
  geom_rect(aes(xmin = wm, xmax = w, ymax = Ladder_score, colour = Region, fill = Region)) +
  theme_bw() +
  labs(x = NULL, y = "Ladder Score of Happiness",
       title = "Variable Width Column Chart for World Happiness Score in 2021") +
  theme_ipsum() +
  theme(legend.position="right")
```



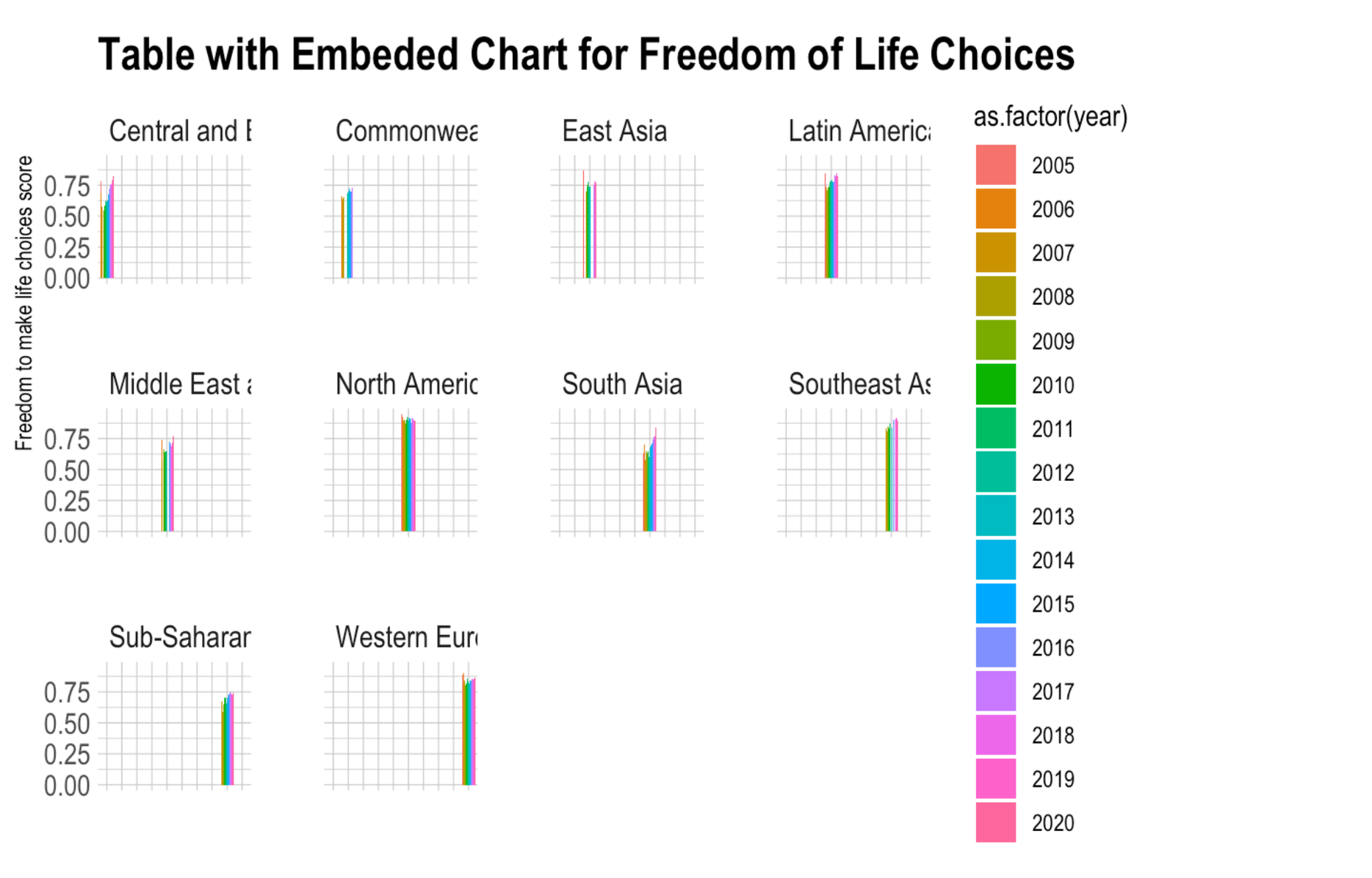

```r
###########
# Sample 2
###########
mydata <- rename(mydata, Population = Population..2020.)

mydata1_2 <- mydata %>% group_by(Region)

Population_by_region <- mydata1_2 %>% summarise(
  Population = mean(Population)
)

Chart1_data2 <- merge(Population_by_region, Number_of_states, by= "Region")
Chart1_data2 <- rename(Chart1_data2, Number_of_states = n)

# Calculate the future positions on the x axis of each bar (left border, central position, right border)
Chart1_data2$w <- cumsum(Chart1_data2$Number_of_states)
Chart1_data2$wm <- Chart1_data2$w - Chart1_data2$Number_of_states
Chart1_data2$wt <- with(Chart1_data2, wm + (w-wm)/2)

# Plot
ggplot(Chart1_data2, aes(ymin = 0)) +
  geom_rect(aes(xmin = wm, xmax = w, ymax = Population, colour = Region, fill = Region)) +
  theme_bw() +
  labs(x = NULL, y = "Average Population by Region",
       title = "Variable Width Column Chart for Average Population in 2020") +
  theme_ipsum() +
  theme(legend.position="right")
```



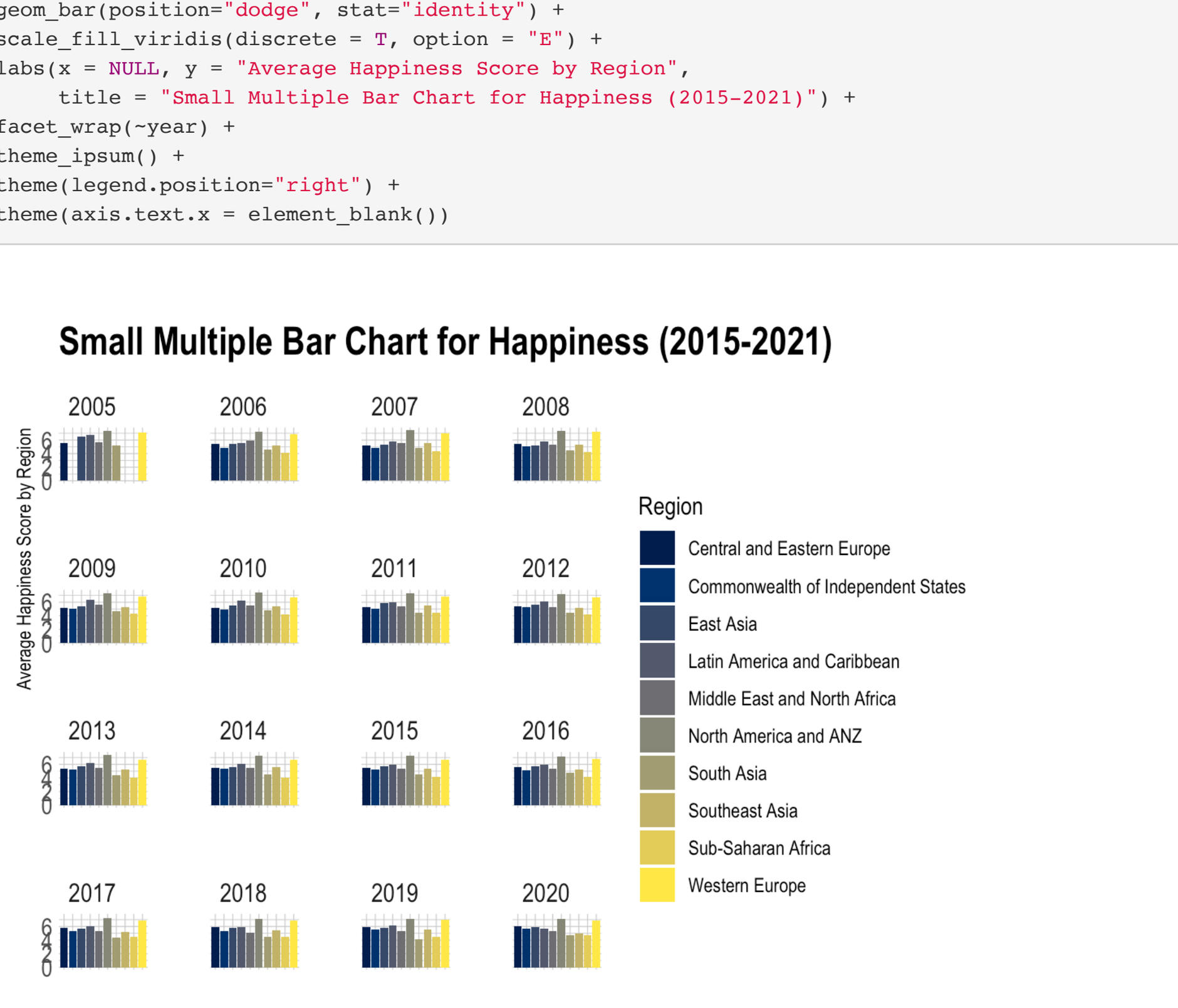

```r
#####################################
# Chart 2 Table with Embeded Chart
#####################################
############
# Sample 1
############
data3 <- read.csv("/Users/min/Desktop/2021 Fall Semester/EPPS 6356 Data Visualization/Assignment 4/world-happines
s-report.csv")
data3 <- rename(data3, Country = Country.name)

mydata2_1 <- subset(data3, year >= 2015)
region_data <- data2[, 1:2]
mydata2_1 <- merge(mydata2_1, region_data, by="Country")
mydata2_1 <- rename(mydata2_1, Region = Regional.indicator)
mydata2_1 <- aggregate(x = mydata2_1$Life.Ladder,
          by = list(mydata2_1$Region, mydata2_1$year),
          FUN=mean)
mydata2_1 <- rename(mydata2_1, Region = Group.1, year = Group.2, Ladder_by_region_year = x)


ggplot(mydata2_1, aes(Region, Ladder_by_region_year, fill=as.factor(year)), angle=45, size=16)+
  geom_bar(position="dodge", stat="identity") + facet_wrap(~Region, nrow=3)+
  labs(x = NULL, y = "Average Happiness Score by Region-year",
       title = "Table with Embeded Chart for Average Happiness Score") +
  theme_ipsum() +
  theme(legend.position="right") +
  theme(axis.text.x = element_blank())
```



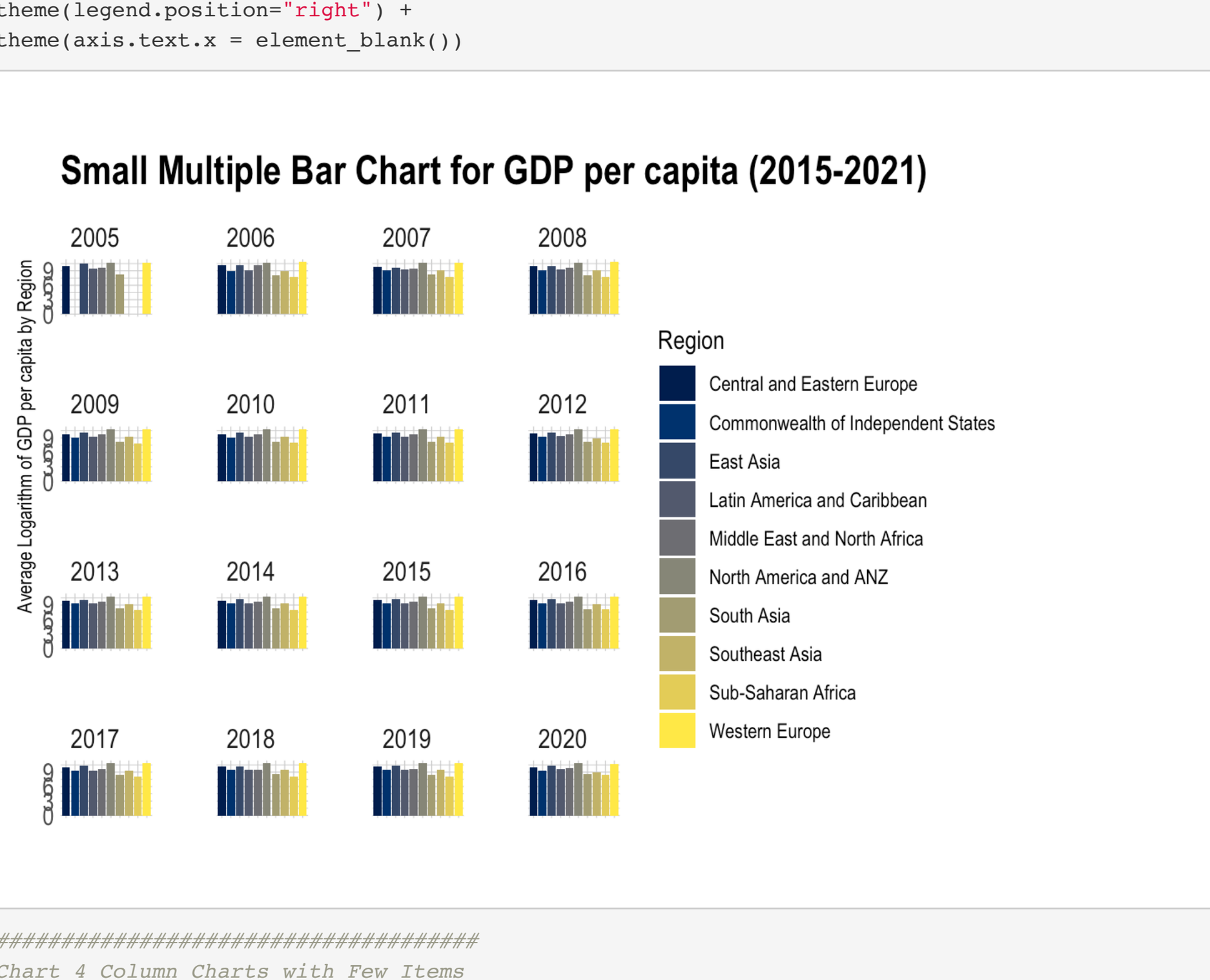

```r
############
# Sample 2
############
mydata2_2 <- merge(data3, region_data, by="Country")
mydata2_2 <- rename(mydata2_2, Region = Regional.indicator)

mydata2_2 <- aggregate(x = mydata2_2$Freedom.to.make.life.choices,
                       by = list(mydata2_2$Region, mydata2_2$year),
                       FUN=mean)
mydata2_2 <- rename(mydata2_2, Region = Group.1, year = Group.2, Freedom.to.make.life.choices = x)


ggplot(mydata2_2, aes(Region, Freedom.to.make.life.choices, fill=as.factor(year)), angle=45, size=16)+
  geom_bar(position="dodge", stat="identity") + facet_wrap(~Region, nrow=3)+
  labs(x = NULL, y = "Freedom to make life choices score",
       title = "Table with Embeded Chart for Freedom of Life Choices") +
  theme_ipsum() +
  theme(legend.position="right") +
  theme(axis.text.x = element_blank())
```



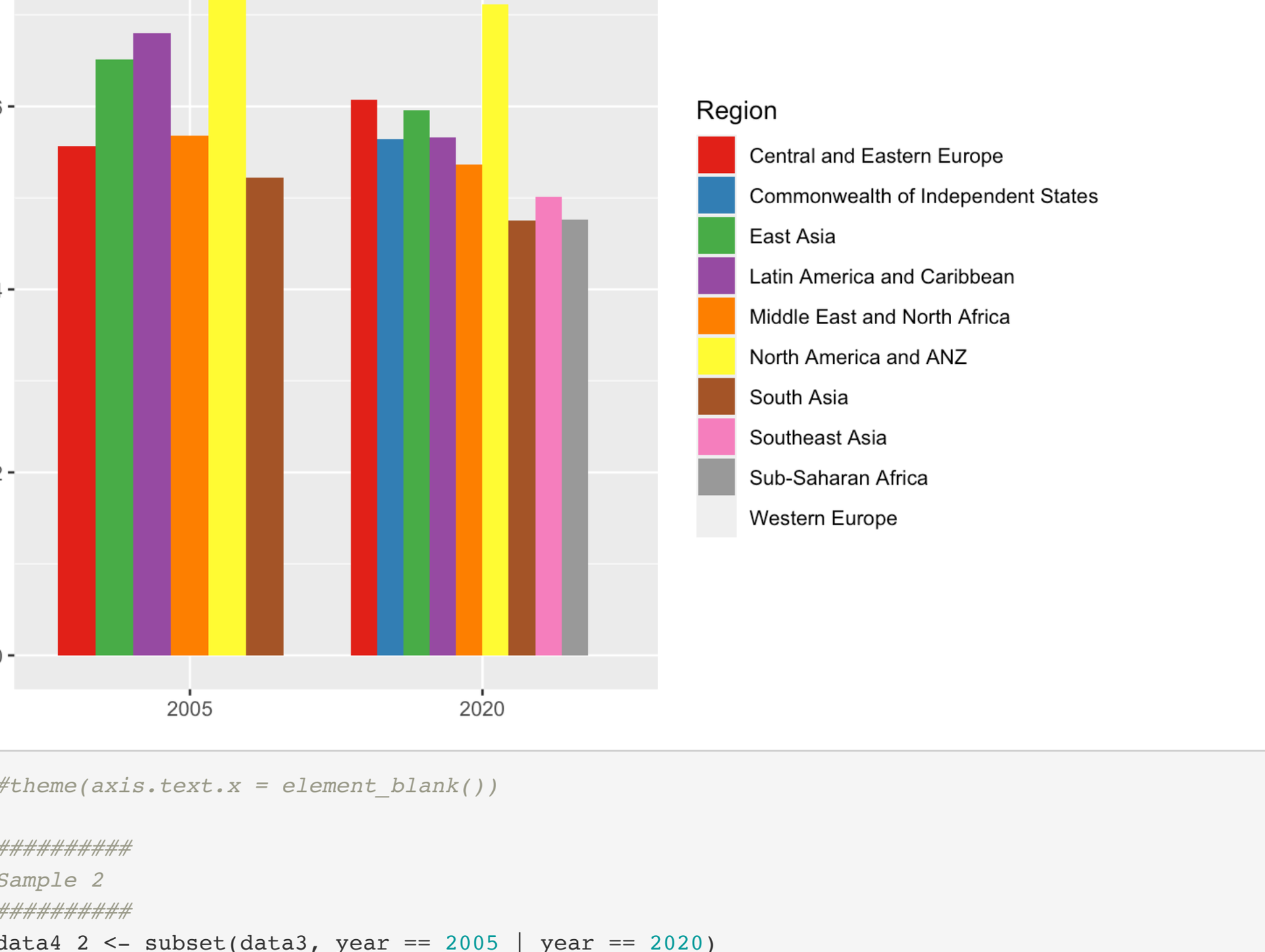

```r
#####################################################
# Chart 3 Bar Charts with Many Items (Small Multiple)
#####################################################
############
# Sample 1
############
mydata3_1 <- merge(data3, region_data, by="Country")
mydata3_1 <- rename(mydata3_1, Region = Regional.indicator)
mydata3_1 <- aggregate(x = mydata3_1$Life.Ladder,
                       by = list(mydata3_1$Region, mydata3_1$year),
                       FUN = mean)
mydata3_1 <- rename(mydata3_1, Region = Group.1, year = Group.2, Ladder_by_region_year = x)

ggplot(mydata3_1, aes(fill = Region, y = Ladder_by_region_year, x = Region)) +
  geom_bar(position="dodge", stat="identity") +
  scale_fill_viridis(discrete = T, option = "E") +
  labs(x = NULL, y = "Average Happiness Score by Region",
       title = "Small Multiple Bar Chart for Happiness (2015-2021)") +
  facet_wrap(~year) +
  theme_ipsum() +
  theme(legend.position="right") +
  theme(axis.text.x = element_blank())
```




```r
############
# Sample 2
############
mydata3_2 <- merge(data3, region_data, by="Country")
mydata3_2 <- rename(mydata3_2, Region = Regional.indicator)

## Filling missing values
# country mean imputation
meanvals <- aggregate(mydata3_2$Log.GDP.per.capita, by=list(mydata3_2$Country), FUN="mean", na.rm=TRUE, na.action
=NULL)
colnames(meanvals) <- c("Country", "ave_log_GDP_per_capita")
mydata3_2 <- merge(x = mydata3_2, y=meanvals, all.x=TRUE, by="Country")
mydata3_2$ave_log_GDP_per_capita[which(!is.na(mydata3_2$Log.GDP.per.capita))] <- mydata3_2$Log.GDP.per.capita[whi
ch(!is.na(mydata3_2$Log.GDP.per.capita))]

# generate the Average Logarithm of GDP per capita by Region
mydata3_2 <- aggregate(x = mydata3_2$ave_log_GDP_per_capita,
                       by = list(mydata3_2$Region, mydata3_2$year),
                       FUN = mean, na.rm=TRUE, na.action=NULL)
mydata3_2 <- rename(mydata3_2, Region = Group.1, year = Group.2, Ave.Log.GDP.per.capita = x)

ggplot(mydata3_2, aes(fill = Region, y = Ave.Log.GDP.per.capita, x = Region)) +
  geom_bar(position="dodge", stat="identity") +
  scale_fill_viridis(discrete = T, option = "E") +
  labs(x = NULL, y = "Average Logarithm of GDP per capita by Region",
       title = "Small Multiple Bar Chart for GDP per capita (2015-2021)") +
  facet_wrap(~year) +
  theme_ipsum() +
  theme(legend.position="right") +
  theme(axis.text.x = element_blank())
```




```r
#########################################
# Chart 4 Column Charts with Few Items
#########################################
############
# Sample 1
############
mydata4_1 <- subset(data3, year == 2005 | year == 2020)
region_data <- data2[, 1:2]
mydata4_1 <- merge(mydata4_1, region_data, by="Country")
mydata4_1 <- rename(mydata4_1, Region = Regional.indicator)
mydata4_1 <- aggregate(x = mydata4_1$Life.Ladder,
                       by = list(mydata4_1$Region, mydata4_1$year),
                       FUN="mean", na.rm=TRUE, na.action=NULL)
mydata4_1 <- rename(mydata4_1, Region = Group.1, year = Group.2, Ladder_by_region_year = x)

ggplot(data = mydata4_1, aes(factor(year), y = Ladder_by_region_year, fill = Region)) +
  geom_bar(position="dodge", stat="identity")+
  labs(x = NULL, y = "Average Happpiness Score by Region",
       title = "Grouped Bar Chart for Happiness in 2005 & 2020") +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position="right")
```




```r
  #theme(axis.text.x = element_blank())

############
# Sample 2
############
mydata4_2 <- subset(data3, year == 2005 | year == 2020)
region_data <- data2[, 1:2]
mydata4_2 <- merge(mydata4_2, region_data, by="Country")
mydata4_2 <- rename(mydata4_2, Region = Regional.indicator)
mydata4_2 <- aggregate(x = mydata4_2$Social.support,
                       by = list(mydata4_2$Region, mydata4_2$year),
                       FUN="mean", na.rm=TRUE, na.action=NULL)
mydata4_2 <- rename(mydata4_2, Region = Group.1, year = Group.2, Social_suppor_by_region_year = x)

ggplot(data = mydata4_2, aes(factor(year), y = Social_suppor_by_region_year, fill = Region)) +
  geom_bar(position="dodge", stat="identity")+
  labs(x = NULL, y = "Average Social Support Score by Region",
       title = "Grouped Bar Chart for Social Support in 2005 & 2020") +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position="right")
```