Data Visualization-03 ggplot2 Differen Chart Types

Created by H. M. Samadhi Chathuranga Rathnayake

```r
setwd("D:\\Workshops\\R Programming for Data Science Workshop\\Part 03 - Data
Visualization\\Datasets")
data=read.csv("gapminder.CSV")
head(data)

#install.packages("ggplot2")
library(ggplot2)

#Different chart types

#Numerical data

#Histogram
ggplot(data,aes(x=gdpPercap))+geom_histogram(bins = 20)

ggplot(data,aes(x=gdpPercap))+geom_density()

#Boxplot
ggplot(data,aes(x=gdpPercap))+geom_boxplot()

#Line plot
ggplot(data,aes(x=as.numeric(row.names(data)),y=gdpPercap))+geom_line()

ggplot(data,aes(x=as.numeric(row.names(data)),y=gdpPercap))+geom_area()

#Categorical data

#Bar charts
ggplot(data,aes(x=continent))+geom_bar()

ggplot(data,aes(y=continent))+geom_bar()

ggplot(data,aes(x=continent,fill=continent))+geom_bar()

#Another type of bar charts
df_cont=data.frame(table(data$continent))
colnames(df_cont)=c("Continent","Frequency")
ggplot(df_cont,aes(x="Continent",y=Frequency, fill=Continent))+
  geom_bar(stat = "identity")

#Pie charts
df_cont=data.frame(table(data$continent))
colnames(df_cont)=c("Continent","Frequency")
ggplot(df_cont,aes(x="Continent",y=Frequency, fill=Continent))+
```

```r
  geom_bar(stat = "identity")+
  coord_polar("y", start=0)

#Numerical VS Numerical

#Scatter plots
ggplot(data, aes(x = gdpPercap, y = lifeExp))+geom_point()

ggplot(data, aes(x = gdpPercap, y = lifeExp))+
  geom_point(alpha = 0.5, col = "red", size = 2)

#Change the colour with another categorical variable
ggplot(data, aes(x = gdpPercap, y = lifeExp, color=continent))+
  geom_point(alpha = 0.5, size = 2)

#Change the size with another numerical variable
ggplot(data, aes(x = gdpPercap, y = lifeExp, color=continent, size=pop))+
  geom_point(alpha = 0.5)

#Categorical VS Categorical

#Multiple bar charts
specie=c(rep("sorgho" , 3) , rep("poacee" , 3) , rep("banana" , 3) ,
         rep("triticum" , 3) )
condition=rep(c("normal" , "stress" , "Nitrogen") , 4)
value=abs(rnorm(12 , 0 , 15))
df=data.frame(specie,condition,value)

head(df)

ggplot(df, aes(fill=condition, y=value, x=specie)) +
  geom_bar(stat="identity",position="dodge")

df_cont_year=data.frame(table(data$continent,data$year))
colnames(df_cont_year)=c("Continent","Year","Count")
head(df_cont_year)

ggplot(df_cont_year, aes(fill=Continent, y=Count, x=Year)) +
  geom_bar(stat="identity",position="dodge")

#Stacked bar charts
specie=c(rep("sorgho" , 3) , rep("poacee" , 3) , rep("banana" , 3) ,
         rep("triticum" , 3) )
condition=rep(c("normal" , "stress" , "Nitrogen") , 4)
value=abs(rnorm(12 , 0 , 15))
df=data.frame(specie,condition,value)

head(df)

ggplot(df, aes(fill=condition, y=value, x=specie)) +
  geom_bar(stat="identity")
```

```r
df_cont_year=data.frame(table(data$continent,data$year))
colnames(df_cont_year)=c("Continent","Year","Count")
head(df_cont_year)

ggplot(df_cont_year, aes(fill=Continent, y=Count, x=Year)) +
  geom_bar(stat="identity")

#100% stacked bar charts
specie=c(rep("sorgho" , 3) , rep("poacee" , 3) , rep("banana" , 3) ,
         rep("triticum" , 3) )
condition=rep(c("normal" , "stress" , "Nitrogen") , 4)
value=abs(rnorm(12 , 0 , 15))
df=data.frame(specie,condition,value)

head(df)

ggplot(df, aes(fill=condition, y=value, x=specie)) +
  geom_bar(stat="identity",position="fill")

df_cont_year=data.frame(table(data$continent,data$year))
colnames(df_cont_year)=c("Continent","Year","Count")
head(df_cont_year)

ggplot(df_cont_year, aes(fill=Continent, y=Count, x=Year)) +
  geom_bar(stat="identity",position="fill")

#Numerical VS Categorical

#Side by side boxplots
ggplot(data,aes(y=gdpPercap,x=continent,fill=continent))+geom_boxplot()

#Other important charts

#Violin charts
ggplot(data,aes(x=continent,y=lifeExp,fill=continent))+geom_violin()

#Heat maps

ggplot(data, aes(x=continent,y=year,fill=gdpPercap))+geom_tile()

ggplot(data, aes(x=continent,y=year,fill=gdpPercap))+geom_tile()+
  scale_fill_distiller(palette = "RdPu")

library(MASS)
data("Boston")
cor_mat=cor(Boston)
cor_mat

df_cor=data.frame(as.table(cor_mat))
colnames(df_cor)=c("V1","V2","Correlation")
ggplot(df_cor, aes(x=V1,y=V2,fill=Correlation))+geom_tile()
```

```
library(ggcorrplot)

## Warning: package 'ggcorrplot' was built under R version 4.0.4

ggcorrplot(cor_mat)

#Pair plots
library(GGally)

head(data)

data_num=data[c("pop","lifeExp","gdpPercap")]

ggpairs(data_num)

data_new=data[c("pop","lifeExp","gdpPercap","continent")]
ggpairs(data_new,aes(colour = continent, alpha = 0.4))
```