Biostatistics: Theory and Applications in R (Virtual)

Week6_Session2_R_training6

Contents

Prof Dr Mohammed Abu Sayed Arfin Khan
Department of Forestry and Environmental Science
Shahjalal University of Science and Technology, Sylhet
+8801917174537, khan-for@sust.edu, nobelarfin@yahoo.com
Homepage | Google scholar | Researchgate | ORCID | Publons | BayCEER

```
##################################
```

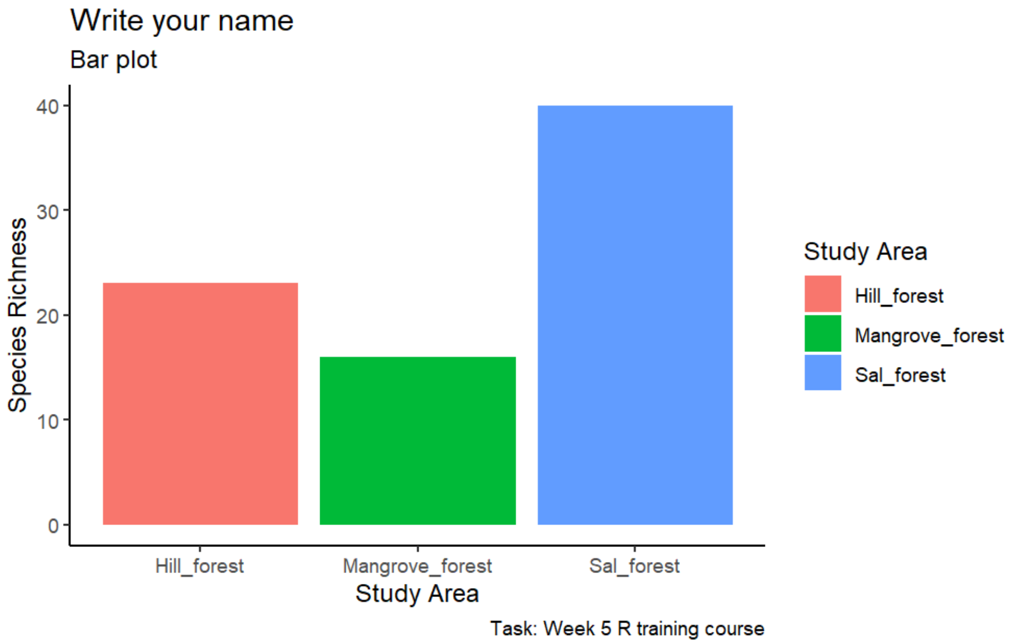
#Task week 5#

```
#use the data set study_2 to create a bar graph
#plot study_area in the x axis and species_richness in the y axis
#use position="dodge" and stat="identity"
#Use legend title "Study Area"
#Use plot title="Write your name",
#Use plot subtitle="Bar plot",
#use y axis name ="Species Richness",
#Use x axis name ="Study Area",
#Use caption as "Task: Week 5 R training course"
#give a name t2 to the bar graph
#add the following theme to the bar graph t2
#theme_excel()
#theme_excel_new()
#theme_few()
#theme_foundation()
#theme_clean()
#theme_igray()
#Create a pdf file with all the bar graph and submit the task

#Set the working directory- getwd()/ setwd("Y:/")
getwd()
setwd("C:/Users/Fahmida Sultana/Desktop/R training/R_training_Class_05")


#install openxlsx package or xlsx package
library(openxlsx)
library(readxl)


#####import data set from xlsx file
study_1 <- read.xlsx("Tree_height.xlsx")
str(study_1)

study_2 <- read_excel("Tree_height.xlsx", sheet = "study_2")
str(study_2)
```

#Week 5 Task bar graph script

```
library(ggplot2)
str(study_2)
```

```
t2 <- ggplot(study_2, aes(x=study_area, y=species_richness, fill=study_area)) +
  geom_bar(position="dodge", stat="identity") +
  guides(fill = guide_legend(title = "Study Area"))+
  labs(title="Write your name",
       subtitle="Bar plot",
       y="Species Richness",
       x="Study Area",
       caption = "Task: Week 5 R training course")
t2
```



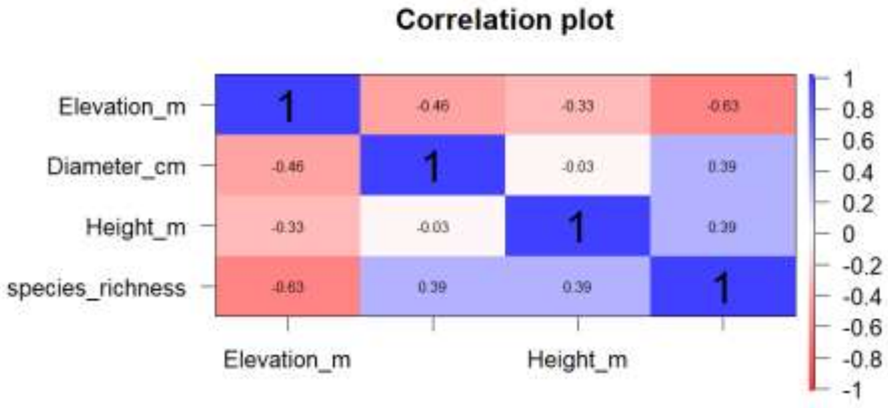
```
#add the following theme to a bar graph
library(ggthemes)
library(hrbrthemes)

t2+theme_excel()
t2+theme_excel_new()
t2+theme_few()
t2+theme_foundation()
t2+theme_clean()
t2+theme_igray()
```

#Correlation Plot function from psych package

```
library(psych)

str(study_2)
```

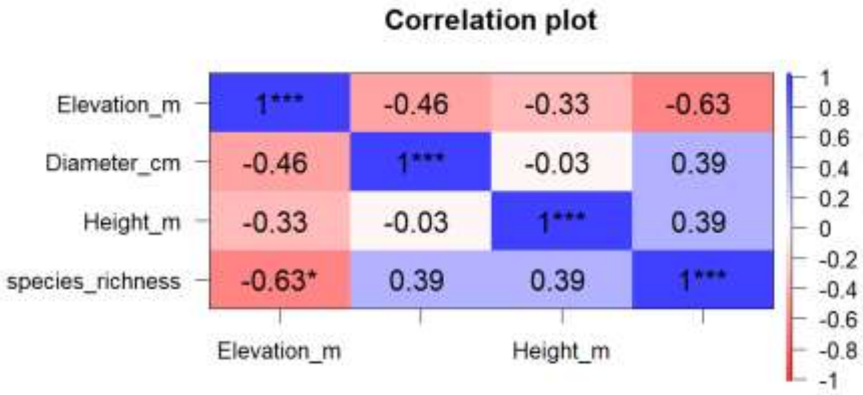
#Correlation text size

```
corPlot(study_2[, 3:6],
        cex = 1.25)
```



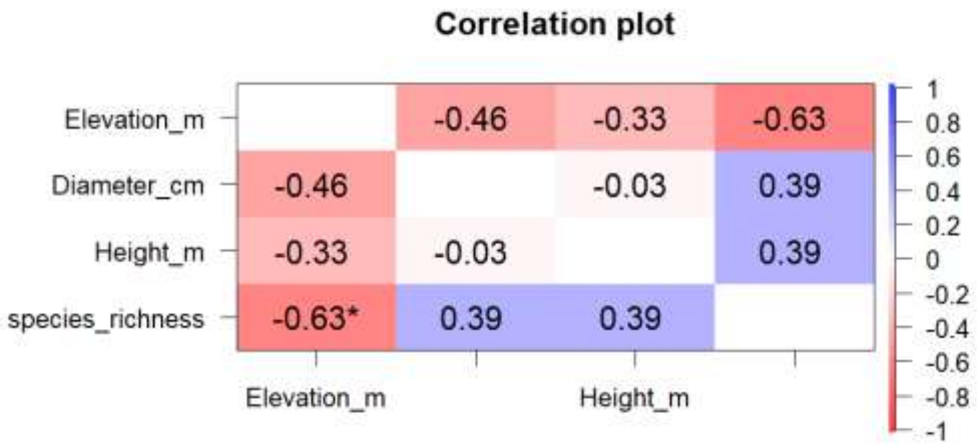
#Show significance levels with astricks setting stars = TRUE

```
corPlot(study_2[, 3:6],
        stars = TRUE, cex = 1.25) #Show significance levels with astricks setting stars = TRUE

corPlot(study_2[, 3:6],
        pval = TRUE, cex = 1.25) #Correlation scaled to p-values
```



#Remove the diagonal

```
corPlot(study_2[, 3:6],
        stars = TRUE, cex = 1.25, diag = FALSE) #Show significance levels with astricks setting stars = TRUE

corPlot(study_2[, 3:6],
        pval = TRUE, cex = 1.25, diag = FALSE) ##Correlation scaled to p-values
```



#Most basic bubble chart with geom_point()

```
library(ggplot2)
library(dplyr)


ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Height_m)) +
  geom_point(alpha=0.8)

ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Diameter_cm)) +
  geom_point(alpha=0.5)
```

#Control circle size with scale_size()

```
ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Height_m)) +
  geom_point(alpha=0.8)+
  scale_size(range = c(.1, 10), name="Height (m)")

ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Diameter_cm)) +
  geom_point(alpha=0.8)+
  scale_size(range = c(.1, 10), name="Diameter (cm)")
```

#Add a fourth dimension: color

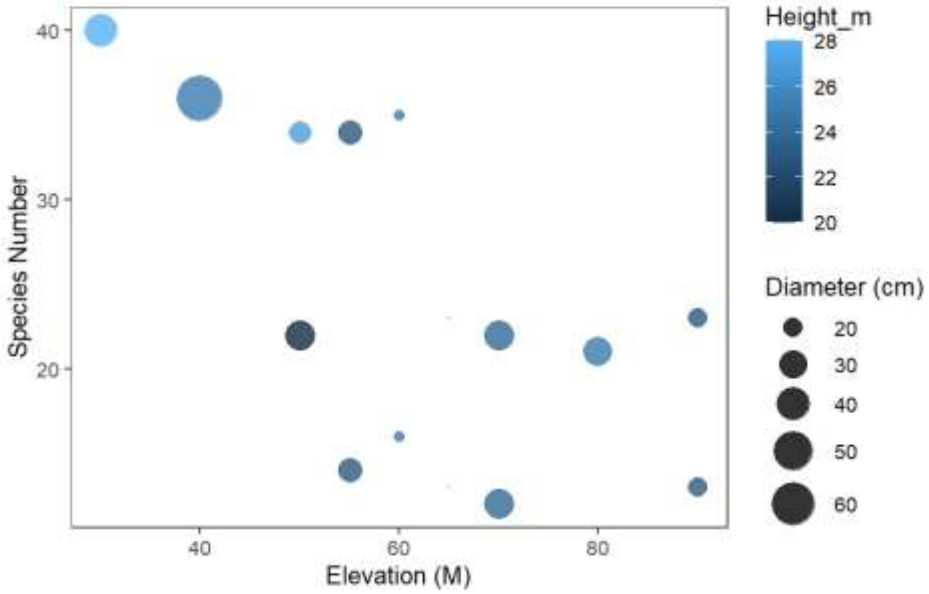
```
ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Diameter_cm, color=Height_m)) +
  geom_point(alpha=0.8)+
  scale_size(range = c(.1, 10), name="Diameter (cm)")
```

#Change color palette, add theme and axis title

```
#use of the viridis package for nice color palette
#use of theme_clean() of the ggthemes package
#custom axis titles with xlab and ylab

library(ggplot2)
library(dplyr)
library(viridis)
library(viridisLite)
library(ggthemes)

ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Diameter_cm, color=Height_m)) +
  geom_point(alpha=0.8)+
  scale_size(range = c(.1, 10), name="Diameter (cm)")+
  scale_fill_viridis(discrete=TRUE, guide="none", option="A") +
  theme_clean() +
  theme(legend.position="right") +
  ylab("Species Number") +
  xlab("Elevation (M)")
```

#use theme_few()

```
ggplot(study_2, aes(x=Elevation_m, y=species_richness, size = Diameter_cm, color=Height_m)) +
  geom_point(alpha=0.8)+
  scale_size(range = c(.1, 10), name="Diameter (cm)")+
  scale_fill_viridis(discrete=TRUE, guide="none", option="A") +
  theme_few() +
  theme(legend.position="right") +
  ylab("Species Number") +
  xlab("Elevation (M)")
```



#2d Histogram with geom_bin2d()

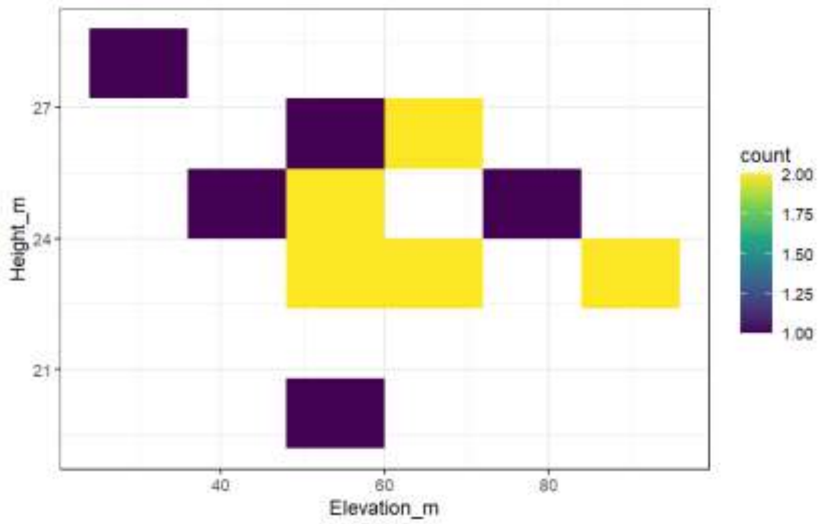
```
str(study_2)

# Basic scatterplot
ggplot(study_2, aes(x=Elevation_m, y=Height_m) ) +
  geom_point()
```

# 2d histogram with default option

```
ggplot(study_2, aes(x=Elevation_m, y=Height_m) ) +
  geom_bin2d() +
  theme_bw()
```

# Bin size control + color palette

```
ggplot(study_2, aes(x=Elevation_m, y=Height_m) ) +
  geom_bin2d(bins = 5) +
  scale_fill_continuous(type = "viridis") +
  theme_bw()
```



#Plot a density chart instead of a histogram

# Show the contour only

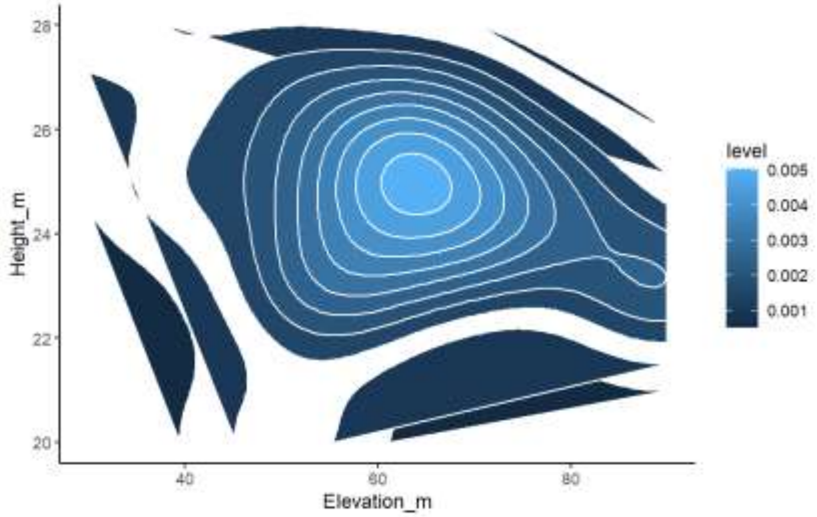
```
ggplot(study_2, aes(x=Elevation_m, y=Height_m) ) +
  geom_density_2d()
```

# Show the area only

```
ggplot(study_2, aes(x=Elevation_m, y=Height_m) ) +
  stat_density_2d(aes(fill = ..level..), geom = "polygon")
```

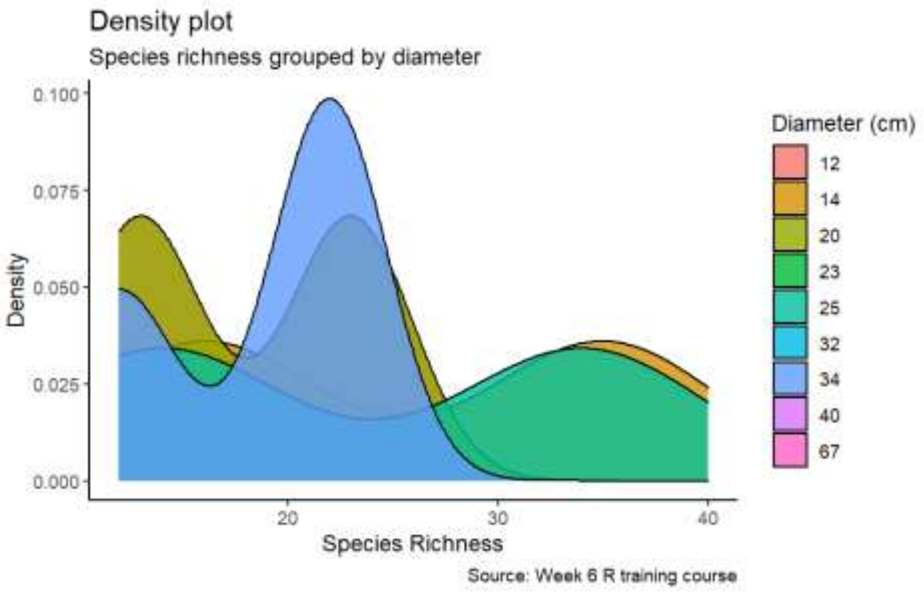
# Area + contour

```
ggplot(study_2, aes(x=Elevation_m, y=Height_m) ) +
  stat_density_2d(aes(fill = ..level..), geom = "polygon", colour="white")
```



#Density plot

```
library(ggplot2)


str(study_2)

# Plot
ggplot(study_2, aes(species_richness))+
   geom_density(aes(fill=factor(Diameter_cm)), alpha=0.8) +
   labs(title="Density plot",
      subtitle="Species richness grouped by diameter",
      caption="Source: Week 6 R training course",
      x="Species Richness",
      y="Density",
      fill="Diameter (cm)")+
      theme_classic()
```



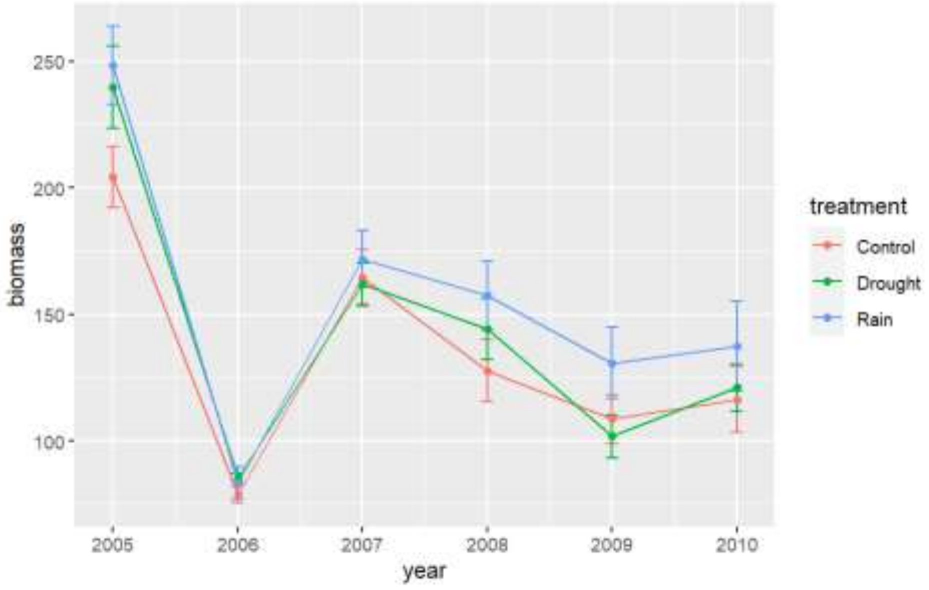
#Summarize the data with function summarySE

```
# With a data set with NA's, use na.rm=TRUE

library(Rmisc)

str(study_1)

study_3 <- summarySE(study_1, measurevar="biomass", groupvars=c("treatment","year"),
na.rm=FALSE)
```

```
study_3
```

#Connected Scatterplot

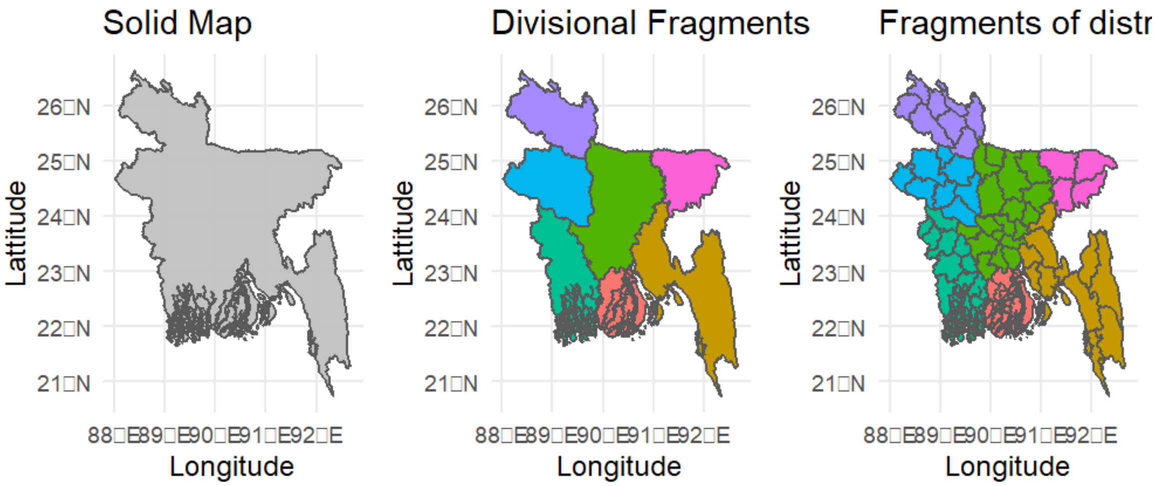
```
library(ggplot2)

# Standard error of the mean
ggplot(study_3, aes(x=year, y=biomass, colour=treatment)) +
  geom_errorbar(aes(ymin=biomass-se, ymax=biomass+se), width=.1) +
  geom_line() +
  geom_point()+
  theme_grey()
```



#Create map of Bangladesh

```
library(ggplot2)
library(sf)
library(gridExtra)

setwd("C:/Users/Fahmida Sultana/Desktop/R training/R_training_Class_06/BD_map")

bd_data0 <- read_sf("BGD_adm0.shp")
bd_data1<- read_sf("BGD_adm1.shp")
bd_data2<- read_sf("BGD_adm2.shp")
```

```
theme_set(theme_minimal())

map0 <- ggplot(bd_data0)+
  geom_sf(fill="grey",alpha=.9)+
  labs(title="Solid Map",x="Longitude",y="Lattitude")

map0

map1 <- ggplot(bd_data1)+
  geom_sf(aes(fill=NAME_1))+
  labs(title="Divisional Fragments",x="Longitude",y="Lattitude")+
  theme(legend.position = "None")

map1

map2 <- ggplot(bd_data2)+
  geom_sf(aes(fill=NAME_1))+
  labs(title="Fragments of districts",
       x="Longitude",y="Lattitude")+
  theme(legend.position = "None")

map2

grid.arrange(map0,map1,map2,nrow=1)
```


Solid Map
Divisional Fragments
Fragments of distr
Lattitude
26□N
25□N
24□N
23□N
22□N
21□N
88□E89□E90□E91□E92□E
Longitude