R Homework

Heva, Prita, Supria

March 29, 2018

Data Visualization in R : Long-term monitoring of a rodent community

Our homework is analyzing database of a long term monitoring for rodent community in Chihuahuan Desert ecosystem near Portal, Arizona, from 1977 to 2000. At this site, 24 experimental plots were established in 1977 and divided among controls and experimental manipulations. The long-term data for the rodent community at the Portal Project has been used to address a variety of questions including

- Monitoring the population-level dynamics of desert rodents & competitive interactions among rodent species.
- Responses of rodents to climatic variability.
- The long-term stability and dynamics of a desert rodent community.

This is our R-homework documentation. The process of data visualisation using R can be divided into four steps :

- Load the library
- Read the data file
- Clean the data
- Time series data of sex and number of sample per plot type
- Correlation between length of hindfoot and weight of animal
- Changes of weight over the year based on each plot type
- Changes of length of hindfoot over the year based on each plot type
- Correlation between hindfoot length and genus
- Correlation between hindfoot length and genus

Load the library

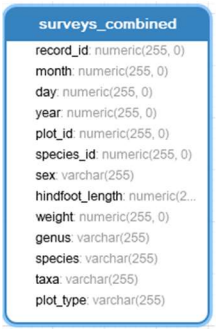
```
library(tidyverse)
library(lubridate)
library(gridExtra)
library(ggplot2)
library(dbplyr)
library(ggpubr)
```

Read the data file

Our team decide to read the combined.csv file because it has the most comprehensive, consise, and compact data.

```
surveys_combined <- read.csv("data/combined.csv")
```

Below is the information about the data structure:



Variable	Levels
species_id	BA,DM,DO,DS,NL,OL,OT,OX,PB,PE,PF,PH,PI,PL,PM,PP,PX,RF,RM,RO,RX,SF,SH,SO
sex	F,M
genus	Baiomys,Chaetodipus,Dipodomys,Neotoma,Onychomys,Perognathus,Peromyscus,Reithrodontomys,Sigmodon
species	albigula,baileyi,eremicus,flavus,fulvescens,fulviventer,hispidus,intermedius,leucogaster,leucopus,maniculatus megalotis,merriami,montanus,ochrognathus,ordii,penicillatus,sp.,spectabilis,taylori,torridus
plot_type	Control,Long-term Krat Exclosure,Rodent Exclosure,Short-term Krat Exclosure,Spectab exclosure

Clean the data

Our team read the raw data and transform it into consistent data that can be analyzed. It is aimed at improving the content of statistical statements based on the data as well as their reliability. This proces is Data Cleaning. In this homework, we ignore the missing data (’’, NULL, is.Na).

```
surveys_combined_clear<- surveys_combined %>% filter(!is.na(sex),
                         sex != "",
                         !is.na(hindfoot_length),
                         hindfoot_length != "",
                         !is.na(weight),
                         weight != "")
```

The result of data cleaning :

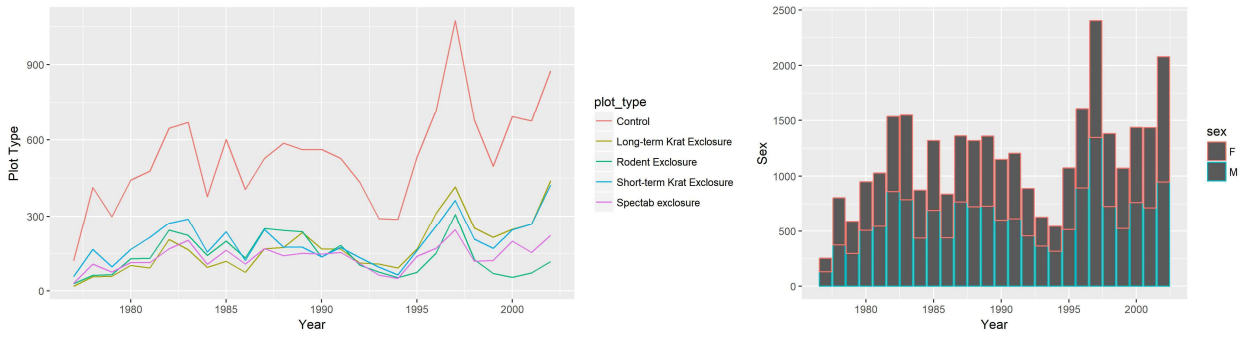
Data Cleaning	
Before	After
34,786	30,676

For the simple distribution tables below the 1st and 3rd Qu. refer to the first and third quartiles, indicating that 25% of the observations have values of that variable which are less than or greater than (respectively) the value listed.

```
     month            day             year         plot_id        species_id   sex       hindfoot_length     weight
 Min.   : 1.000   Min.   : 1.00   Min.   :1977   Min.   : 1.00   DM     :6784   F:10231   Min.   : 6.00   Min.   :  4.00
 1st Qu.: 4.000   1st Qu.: 9.00   1st Qu.:1985   1st Qu.: 5.00   PP     :2104   M:11242   1st Qu.:21.00   1st Qu.: 20.00
 Median : 7.000   Median :16.00   Median :1991   Median :11.00   PB     :1977             Median :31.00   Median : 36.00
 Mean   : 6.551   Mean   :16.04   Mean   :1991   Mean   :11.21   DO     :1943             Mean   :29.19   Mean   : 41.76
 3rd Qu.:10.000   3rd Qu.:23.00   3rd Qu.:1997   3rd Qu.:17.00   RM     :1677             3rd Qu.:36.00   3rd Qu.: 47.00
 Max.   :12.000   Max.   :31.00   Max.   :2002   Max.   :24.00   OT     :1468             Max.   :58.00   Max.   :270.00
                                                                 (Other):5520
          genus             species                        plot_type
 Dipodomys      :10130   merriami    :6784   Control                 :9725
 Chaetodipus    : 4088   penicillatus:2104   Long-term Krat Exclosure :3188
 Onychomys      : 2115   baileyi     :1977   Rodent Exclosure         :2505
 Reithrodontomys: 1732   ordii       :1943   Short-term Krat Exclosure:3572
 Peromyscus     : 1469   megalotis   :1677   Spectab exclosure        :2483
 Perognathus    : 1028   torridus    :1468
 (Other)        :  911   (Other)     :5520
```

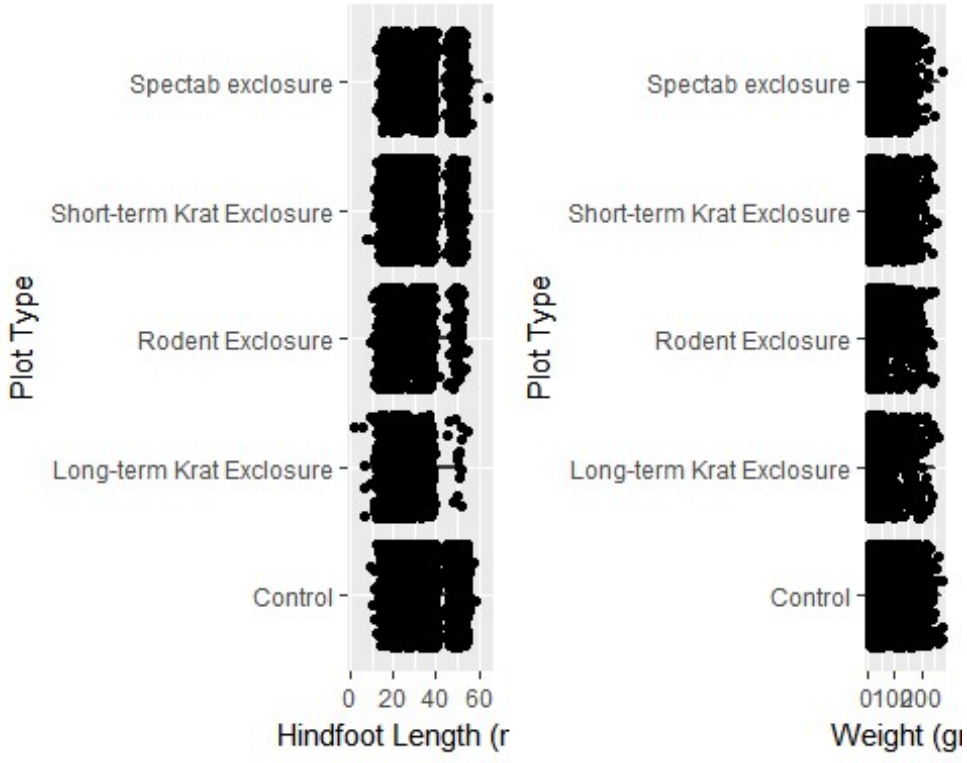
Time series data of sex and number of sample per plot type

```
#create line chart plot type per year
year_plot_type <- surveys_combined_clear %>% group_by(year, plot_type) %>% ta
lly()
line_chart <- ggplot(year_plot_type, aes(x=year, y=n, color=plot_type)) +
  geom_line() + xlab("Year") + ylab("Plot Type")

#create bar chart sex per year
year_sex <- surveys_combined_clear %>% group_by(year, sex) %>% tally()
bar_chart <- ggplot(year_sex, aes(x=year, y=n, color=sex)) +
  geom_bar(stat="identity") + xlab("Year") + ylab("Sex")

#put chart to grid
timeseries_plot <- grid.arrange(line_chart, bar_chart, ncol=2, widths=c(9,6))

#save plot into image
ggsave("image/plot1.jpg", timeseries_plot, width=15, dpi=300)
```



The line graph illustrates the number of rodent sample among controls and other experimental manipulations. Each sample indicates data for every desert rodent caught on the 20 ha. And the bar graph shows the number of rodent differentiate by sex. Overall, there is a trend of decreasing number of sample but it increase in 2000.

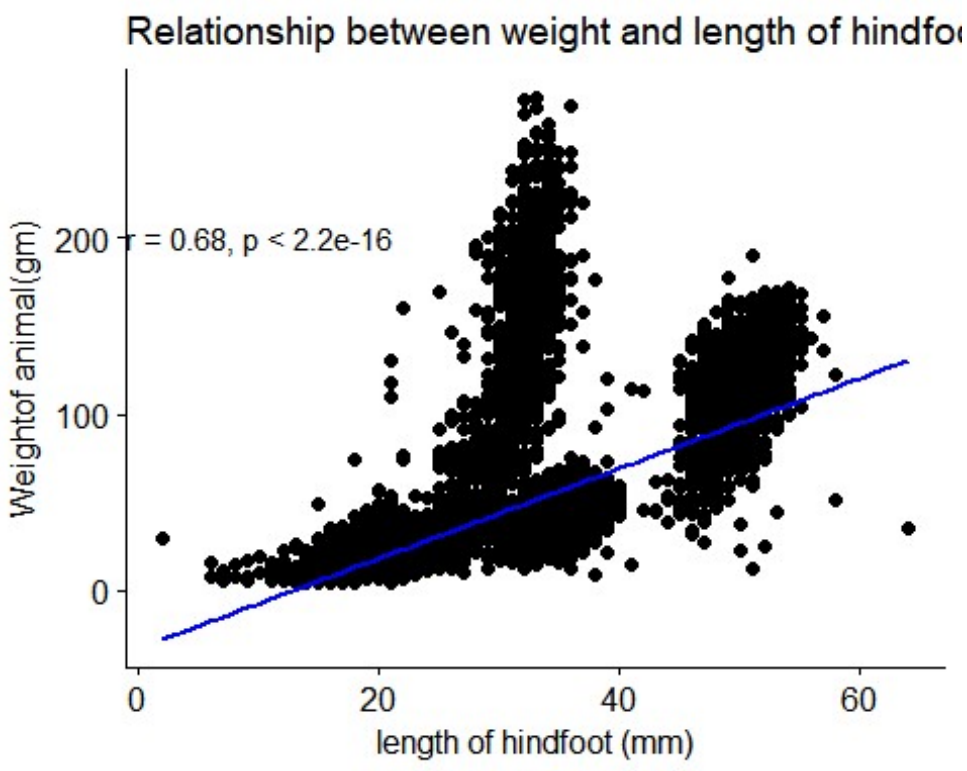
```
#create boxplot chart weight per plot_type
boxplot_chart_weight <- ggplot(surveys_combined_clear, aes(x=weight, y=plot_t
ype))+
  geom_boxplot()+xlab("Weight (gr)")+ylab("Plot Type") +  geom_jitter()

#create boxplot chart hindfoot length per plot_type
boxplot_chart_length <- ggplot(surveys_combined_clear, aes(x=hindfoot_length,
y=plot_type))+
geom_boxplot()+xlab("Hindfoot Length (mm)")+ylab("Plot Type") + geom_jitter()

#put chart to grid
frequency_plot <- grid.arrange(boxplot_chart_length, boxplot_chart_weight, nc
ol=2, widths=c(8,7))

#save plot into image
ggsave("image/plot2.jpg", frequency_plot, width=15, dpi=300)
```



This boxplot graph explain the distribution of hindfoot length and weight per plot type. Detail correlation will be explain below.

Correlation between length of hindfoot and weight of animal

```
# Created by supria
# Scatter plot with correlation coefficient
#:::::::::::::::::::::::::::::::::::::::::::::::
sp <- ggscatter(surveys_combined_clear, x = "hindfoot_length", y = "weight",t
itle = "Relationship between weight and length of hindfoot", xlab = "length o
f hindfoot (mm)", ylab = "Weightof animal(gm)",
                add = "reg.line",  # Add regressin line
                add.params = list(color = "blue", fill = "lightgray"), # Cust
omize reg. line
                conf.int = TRUE # Add confidence interval
)
# Add correlation coefficient
final.plot<- sp + stat_cor(method = "pearson", label.x = 10, label.y = 200)
final.plot
```

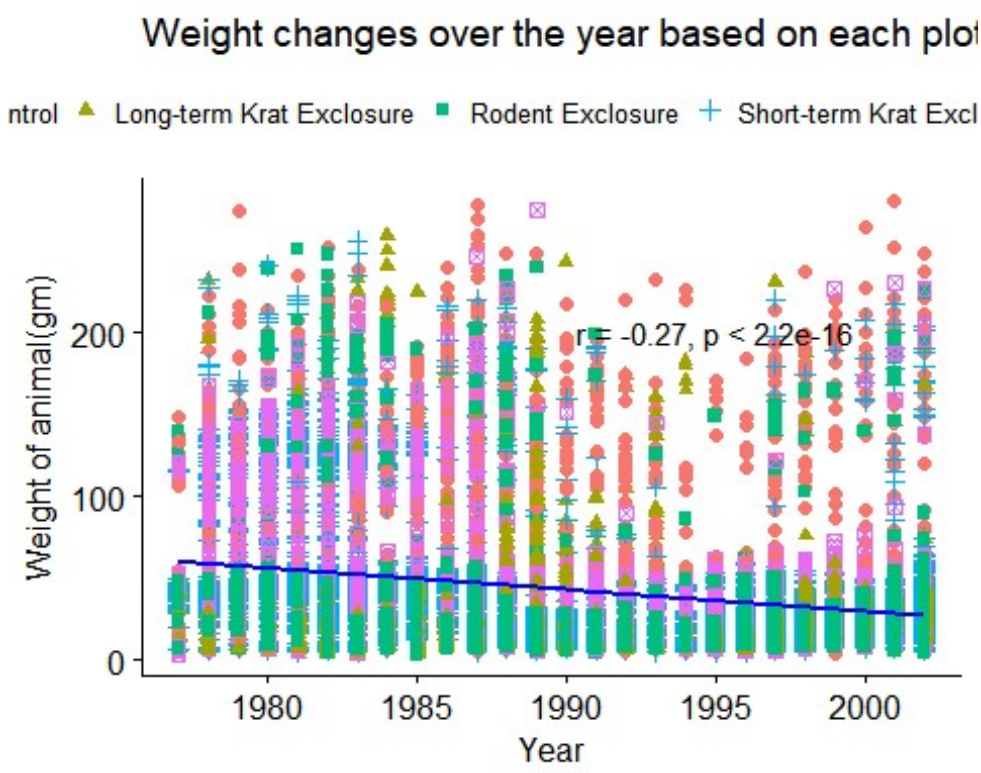



```
#save plot into image
ggsave("image/plot3.jpg", final.plot, width=10, dpi=300)
```

This plot means that what is the relationship between weight of animal and length of hindfoot length. Using statistic analysis we found that thera is linear correlation. And the value of R-squared is greater than .5. It has shown there is 68% linearly correlated.

Changes of weight over the year based on each plot type

```
# created by supria
wg<- ggscatter(data=surveys_combined_clear, x='year',y='weight', color ="plot
_type",shape = "plot_type",
               title = "Weight changes over the year based on each plot type"
, xlab = "Year", ylab = "Weight of animal(gm)",
          add = "reg.line",  # Add regressin line
          add.params = list(color = "blue","red","green","yellow","pink", fil
l = "lightgray"), # Customize reg. line
          conf.int = TRUE # Add confidence interval
 )
# Add correlation coefficient
final.wg <- wg + stat_cor(method = "pearson", label.x = 1995, label.y = 200)
final.wg
```

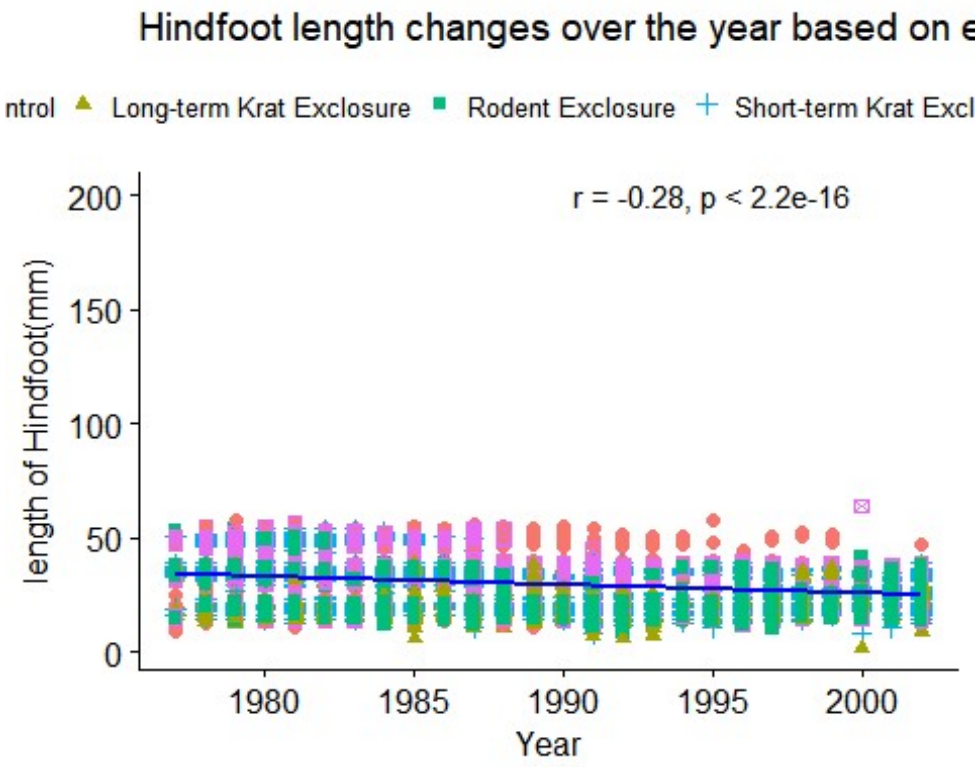



```
ggsave("image/plot4.jpg", final.wg, width=10, dpi=300)
```

This plot explains about what is the change of weight over the year for each plot_type. It has shown that there is no relationship of weight over the year. R-squared value explains that -.27 which is very low.

Changes of length of hindfoot over the year based on each plot type

```
# Created by supria
# Extend the regression lines beyond the domain of the data
##hindfoot changes over the year

hd<- ggscatter(data=surveys_combined_clear, x='year',y='hindfoot_length', col
or ="plot_type",shape = "plot_type",
               title = "Hindfoot length changes over the year based on each p
lot type", xlab = "Year", ylab = "length of Hindfoot(mm)",
               add = "reg.line",  # Add regressin line
               add.params = list(color = "blue","red","green","yellow","pink"
, fill = "lightgray"), # Customize reg. line
               conf.int = TRUE # Add confidence interval
)
# Add correlation coefficient
final.hd<- hd + stat_cor(method = "pearson", label.x = 1995, label.y = 200)
final.hd
```




```
ggsave("image/plot5.jpg", final.wg, width=10, dpi=300)
```

This plot explains about what is the change of length of hindfoot over the year for each plot_type. It has shown that there is no relationship of weight over the year. R-squared value explains that -.28 which is very low.

Heva : Correlation between hindfoot length and genus

```
#heva
```

Heva : Correlation between hindfoot length and genus

```
#heva
```