post01-Yichi-Zhang

Introduction

Motivation

As I proceeded with the study in R language, I realized that in order to communicate information clearly and efficiently, statistical graphics, plots and information graphics have become an integral part of the learning and analyzing process. It is also the fact that the evolution of data visualization has been done by well-known practitioners though out the history. The most famous early visualization data described by Charles Minard appears to be in Napoleon's March, where the data contains the information of the temperature impact on Napoleon's invasion of Russia over time.

Overall Content

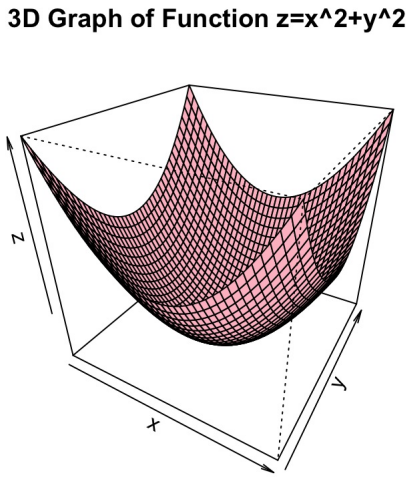
The R language provides a satisfactory set of built-in functions and libraries (such as ggplot2, leaflet, lattice) used to create visual effects to render data. Today, I'll explore more advanced functions for data visualization based on what we've learned in class, such as 3D plots for multivariable functions, Maps and Venn diagrams.

3D plots in R

3D plots for multivariable functions

We've learned how to create a function with one variable and graph it in the two-dimensional space, and now let's explore the possibilities of graphing a function with two variables in a three dimensional space. In order to do so, we'll need to install the package "Lattice". Below is a example of making the 3D graph for a function $z = x^2 + y^2$

```
#install.packages("lattice")
library(lattice)

f <- function(x, y) {
return (x ^ 2 + y ^ 2)
}
x <- seq(-10, 10, length = 40)
y <- x
z <- outer(x, y, f)
z[is.na(z)] <- 1

#3D graph
persp(
x,
y,
z,
theta = 30,
phi = 30,
expand = 1,
col = "pink",
main = "3D Graph of Function z=x^2+y^2"
)
```



#####3D Bar Plot Using 3D bar plot,

we will be able to examine the data frame from another perspective, and here is the example of using NBA player's data to find the relationship between age and position of players in different teams. From this plot we will be able to tell player's age differences between different positions and Teams. In order to do so we will need to install the package "latticeExtra"

```
#importing data
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
dat <- read.csv("../data/nba2017-roster.csv")
df <- dat %>%
filter(team == "WAS" | team == "LAL" | team == "BOS")

slice(df, 1:10)
```

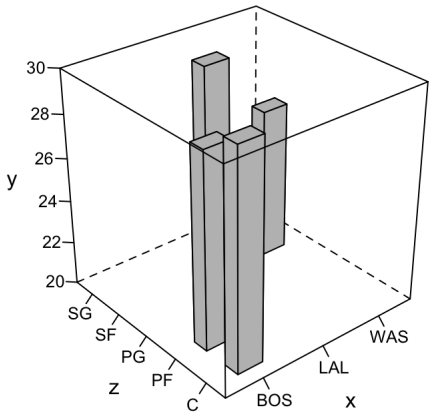
```
## # A tibble: 10 x 8
##              player   team position height weight   age experience
##              <fctr> <fctr>   <fctr>  <int>  <int> <int>      <int>
##  1       Al Horford    BOS        C     82    245    30          9
##  2     Amir Johnson    BOS       PF     81    240    29         11
##  3    Avery Bradley    BOS       SG     74    180    26          6
##  4 Bojan Bogdanovic    WAS       SF     80    216    27          2
##  5     Bradley Beal    WAS       SG     77    207    23          4
##  6   Brandon Ingram    LAL       SF     81    190    19          0
##  7 Brandon Jennings    WAS       PG     73    170    27          7
##  8 Chris McCullough    WAS       PF     83    200    21          1
##  9     Corey Brewer    LAL       SF     81    186    30          9
## 10 D'Angelo Russell    LAL       PG     77    195    20          1
## # ... with 1 more variables: salary <dbl>
```

```
df <- select(df, team, position, age)

#making a table using the data
d <- read.table(text = ' x   y     z
BOS   30   C
BOS   29   PF
WAS   27   SF
WAS   23   SG
LAL   30   SF
LAL   20   PG',
header = TRUE)

library(latticeExtra)
```

```
## Loading required package: RColorBrewer
```

```
cloud(
y ~ x + z,
d,
panel.3d.cloud = panel.3dbars,
col.facet = 'grey',
main = "Players'Age Differences Between Positions and Teams",
xbase = 0.4,
ybase = 0.4,
scales = list(arrows = FALSE, col = 1),
par.settings = list(axis.line = list(col = "transparent"))
)
```

Players'Age Differences Between Positions and Teams



Map in R

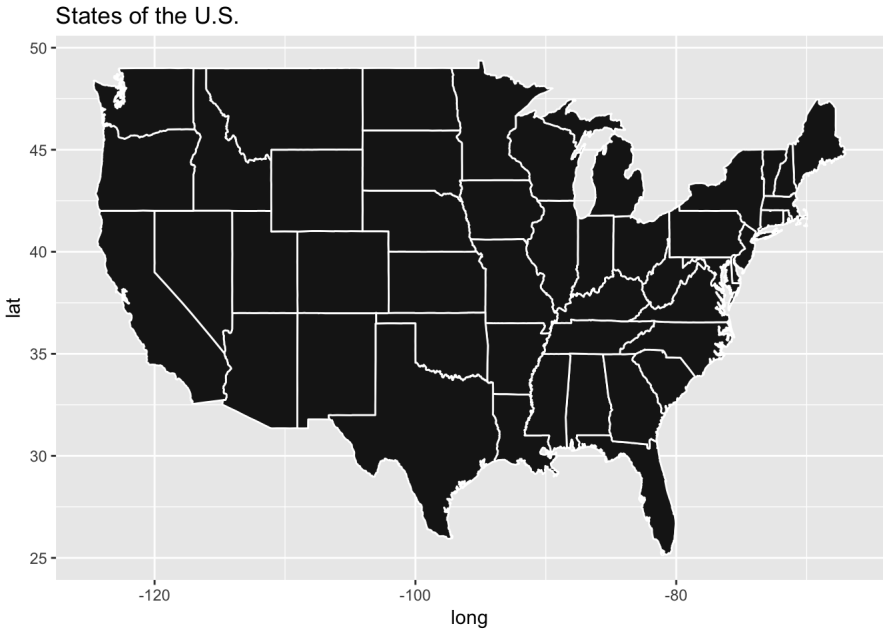
In London in 1854, John Snow, another pioneer in data visualization used map to find that the source of the cholera epidemic was related to the public water pump, and the informational map he created helped to pinpoint the precise location of the root of the burst. Now let's explore the possibilities of using ggplot2 to create various maps. ####Plotting maps-package maps with ggplot The example below shows how to use ggplot to create a map of the states of the U.S

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

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:latticeExtra':
##
##     layer
```

```
library(maps)
#load us map data
all_states <- map_data("state")
#plot all states with ggplot
p <- ggplot()
p <-
p + geom_polygon(
data = all_states,
aes(x = long, y = lat, group = group),
colour = "white",
fill = "grey10"
) +
ggtitle("States of the U.S.")
p
```



Venn diagrams in R

Venn diagrams of two variables

Venn diagrams are a very commonly used graphing technique that illustrates levels of overlap between groups in data. They can be created in R using code written as part of the Bioconductor Project. This is an example of using venn diagram to distinguish dog, cat and lizard people

```
#install.package("VennDiagram")
library(VennDiagram)
```

```
## Loading required package: grid
```

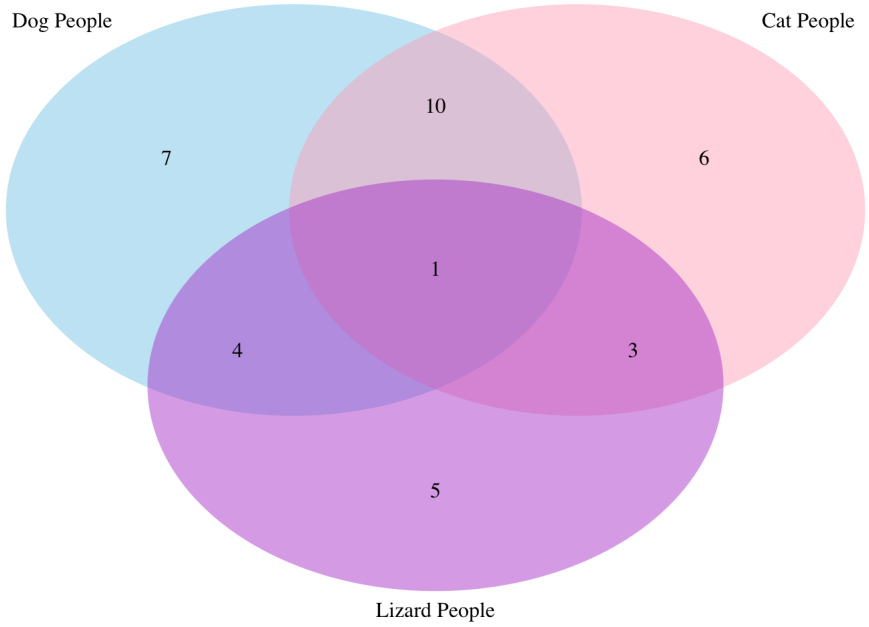
```
## Loading required package: futile.logger
```

```
library(gridExtra)
```

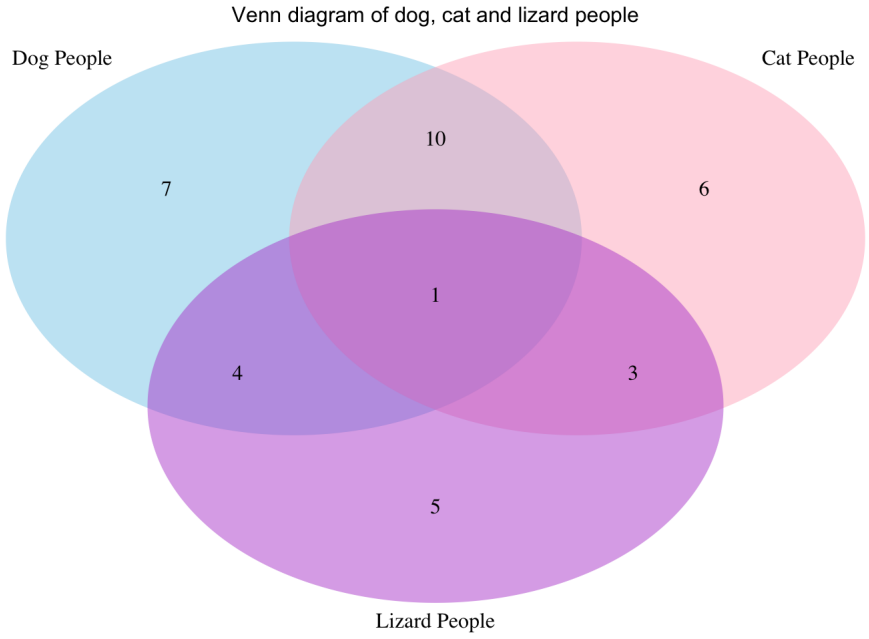
```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(grid)
grid.newpage()
v <-
draw.triple.venn(
area1 = 22,
area2 = 20,
area3 = 13,
n12 = 11,
n23 = 4,
n13 = 5,
n123 = 1,
category = c("Dog People", "Cat People", "Lizard People"),
lty = "blank",
fill = c("skyblue", "pink1", "mediumorchid")
)
```




```
venn <-
grid.arrange(gTree(children = v), top = "Venn diagram of dog, cat and lizard people")
```



Take home messages

Effective visualization helps us analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. We may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic follows the task. It is important for us to pratice more in terms of data visualization in R so that we have a better understanding of our data.

Reference: https://www.programiz.com/r-programming/3d-plot https://www.r-bloggers.com/3d-density-plot-in-r-with-plotly/ https://cran.r-project.org/web/packages/dplyr/dplyr.pdf https://www.statmethods.net/graphs/density.html https://discuss.analyticsvidhya.com/t/how-to-add-[illegible]-to-a-data-frame-in-r/3278 https://stackoverflow.com/questions/26794236/ggplot2-3d-bar-plot https://www.seehuhn.de/blog/123.html