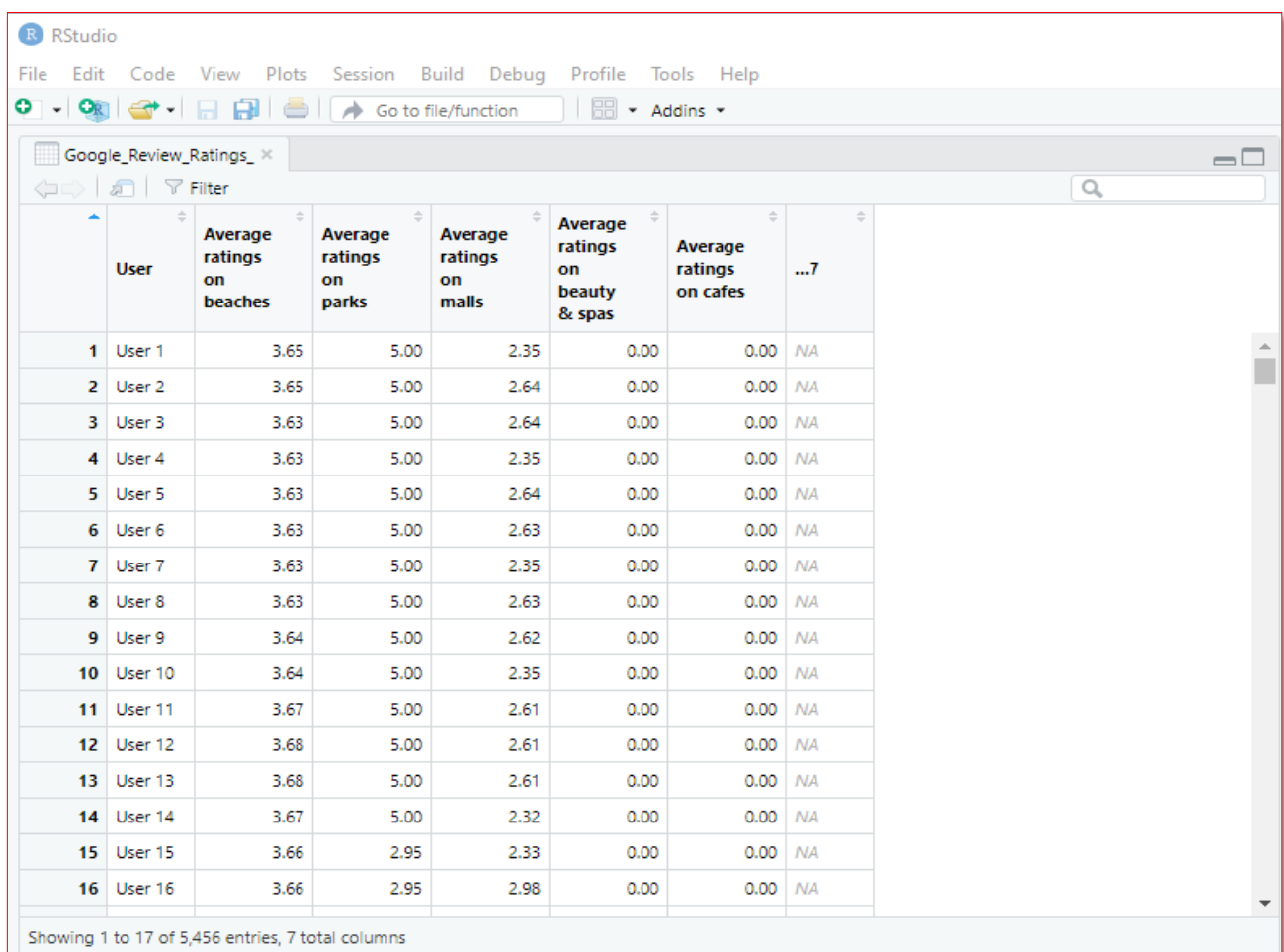


01076253 PROBABILITY AND STATISTICS HW 2

สืบเนื่องมาจาก HW 1 ที่ทางผู้จัดทำได้หยิบยกชุดข้อมูลที่เกี่ยวข้องกับ Travel Review Rating Dataset (From the Machine Learning Repository of University of California) โดยประกอบด้วยคอลัมน์ Numeric (ประเภทการรีวิวแยก 5 ประเภท), และ Category (หมายเลข User ผู้รีวิวสถานที่เหล่านั้น) และเพื่อที่จะได้คำนวณหาค่าสถิติพื้นฐาน และวิเคราะห์ข้อมูลจากกราฟได้ง่ายดายมากขึ้น ทางผู้จัดทำจึงได้ใช้โปรแกรม RStudio ในการวิเคราะห์ข้อมูลด้วย ภาษา R โดยมีรายละเอียดดังต่อไปนี้



	User	Average ratings on beaches	Average ratings on parks	Average ratings on malls	Average ratings on beauty & spas	Average ratings on cafes	...7
1	User 1	3.65	5.00	2.35	0.00	0.00	NA
2	User 2	3.65	5.00	2.64	0.00	0.00	NA
3	User 3	3.63	5.00	2.64	0.00	0.00	NA
4	User 4	3.63	5.00	2.35	0.00	0.00	NA
5	User 5	3.63	5.00	2.64	0.00	0.00	NA
6	User 6	3.63	5.00	2.63	0.00	0.00	NA
7	User 7	3.63	5.00	2.35	0.00	0.00	NA
8	User 8	3.63	5.00	2.63	0.00	0.00	NA
9	User 9	3.64	5.00	2.62	0.00	0.00	NA
10	User 10	3.64	5.00	2.35	0.00	0.00	NA
11	User 11	3.67	5.00	2.61	0.00	0.00	NA
12	User 12	3.68	5.00	2.61	0.00	0.00	NA
13	User 13	3.68	5.00	2.61	0.00	0.00	NA
14	User 14	3.67	5.00	2.32	0.00	0.00	NA
15	User 15	3.66	2.95	2.33	0.00	0.00	NA
16	User 16	3.66	2.95	2.98	0.00	0.00	NA

Showing 1 to 17 of 5,456 entries, 7 total columns

ภาพแสดงตารางข้อมูลเมื่อ import file .csv เข้ามาในโปรแกรม RStudio

ข้อมูลการคำนวณข้อมูลเชิงสถิติพื้นฐานแต่ละคอลัมน์

```

Console Terminal x Jobs x
~/
> view(Google_Review_Ratings_)
> mean(Google_Review_Ratings_.$'Average ratings on beaches')
[1] 2.796886
> which.max(table(Google_Review_Ratings_.$'Average ratings on beaches'))
5
293
> median(Google_Review_Ratings_.$'Average ratings on beaches')
[1] 2.46
> sd(Google_Review_Ratings_.$'Average ratings on beaches')
[1] 1.309159
> Google_Review_Ratings_ >> Average ratings on beaches

```

ข้อมูลการคำนวณข้อมูลเชิงสถิติของคอลัมน์ *Average ratings on beaches*

```

Console Terminal x Jobs x
~/
> view(Google_Review_Ratings_)
> mean(Google_Review_Ratings_.$'Average ratings on parks')
[1] 2.958941
> which.max(table(Google_Review_Ratings_.$'Average ratings on parks'))
5
316
> median(Google_Review_Ratings_.$'Average ratings on parks')
[1] 2.67
> sd(Google_Review_Ratings_.$'Average ratings on parks')
[1] 1.339056
> Google_Review_Ratings_ >>> Average ratings on parks

```

ข้อมูลการคำนวณข้อมูลเชิงสถิติของคอลัมน์ *Average ratings on parks*

```

Console Terminal x Jobs x
~/
> view(Google_Review_Ratings_)
> mean(Google_Review_Ratings_.$'Average ratings on malls')
[1] 2.540795
> which.max(table(Google_Review_Ratings_.$'Average ratings on malls'))
5
334
> median(Google_Review_Ratings_.$'Average ratings on malls')
[1] 2.17
> sd(Google_Review_Ratings_.$'Average ratings on malls')
[1] 1.111391
> Google_Review_Ratings_ >>> Average ratings on malls|

```

ข้อมูลการคำนวณข้อมูลเชิงสถิติของคอลัมน์ *Average ratings on malls*

```

Console Terminal x Jobs x
~/
> view(Google_Review_Ratings_)
> mean(Google_Review_Ratings_.$'Average ratings on beauty & spas')
[1] 0.9658376
> which.max(table(Google_Review_Ratings_.$'Average ratings on beauty & spas'))
0
1
> median(Google_Review_Ratings_.$'Average ratings on beauty & spas')
[1] 0.76
> sd(Google_Review_Ratings_.$'Average ratings on beauty & spas')
[1] 0.9298533
> Google_Review_Ratings_ >>> Average ratings on beauty & spas|

```

ข้อมูลการคำนวณข้อมูลเชิงสถิติของคอลัมน์ *Average ratings on beauty & spas*

```

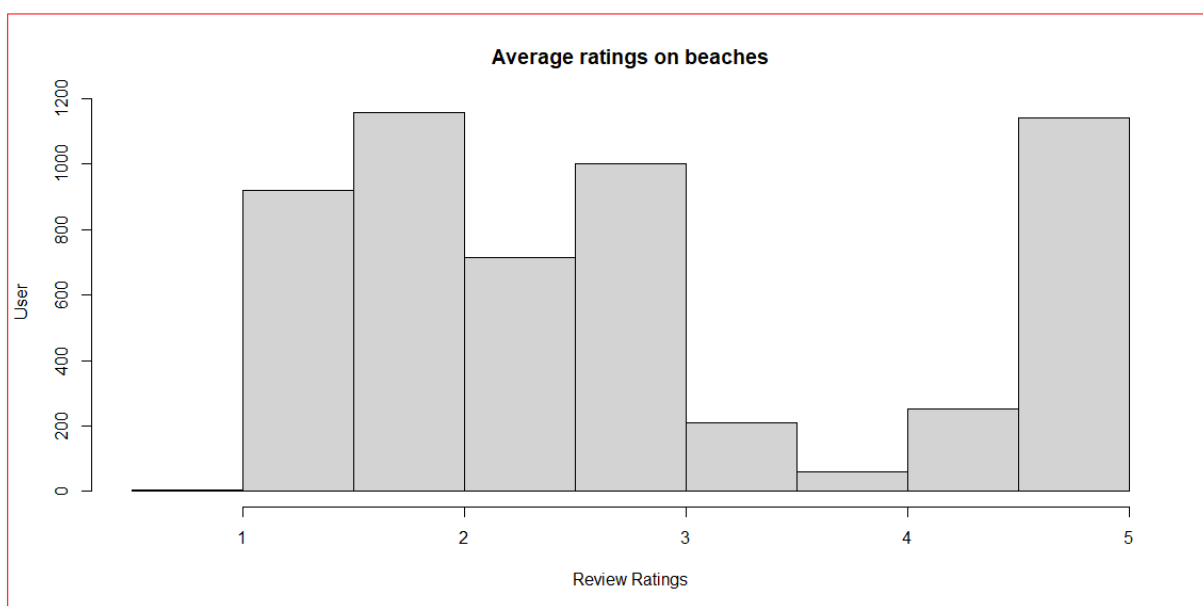
Console Terminal x Jobs x
~/
> view(Google_Review_Ratings_)
> mean(Google_Review_Ratings_.$'Average ratings on cafes')
[1] 1.750537
> which.max(table(Google_Review_Ratings_.$'Average ratings on cafes'))
5
343
> median(Google_Review_Ratings_.$'Average ratings on cafes')
[1] 1.03
> sd(Google_Review_Ratings_.$'Average ratings on cafes')
[1] 1.598734
> Google_Review_Ratings_ >>> Average ratings on cafes

```

ข้อมูลการคำนวณข้อมูลเชิงสถิติของคอลัมน์ *Average ratings on cafes*

[illegible]

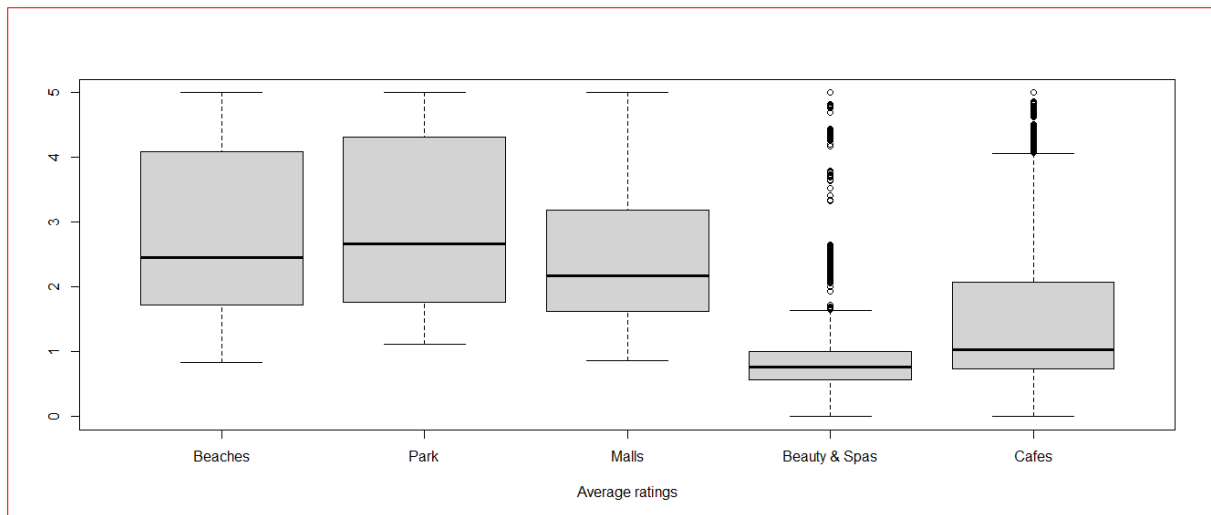
ข้อมูลแสดงกราฟ Stem and leaf ของคอกลิ้ม Average ratings on cafes (ข้อมูลแสดงจำนวนคะแนนเฉลี่ยของการรีวิว จากคะแนนเต็ม 5 คะแนน ของ User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)



Histogram ของคอลัมน์ Average ratings on beaches (ข้อมูลแสดงจำนวนคะแนนเฉลี่ยของการรีวิว จากคะแนนเต็ม 5 คะแนน ของ User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)

Code สำหรับสร้าง Histogram

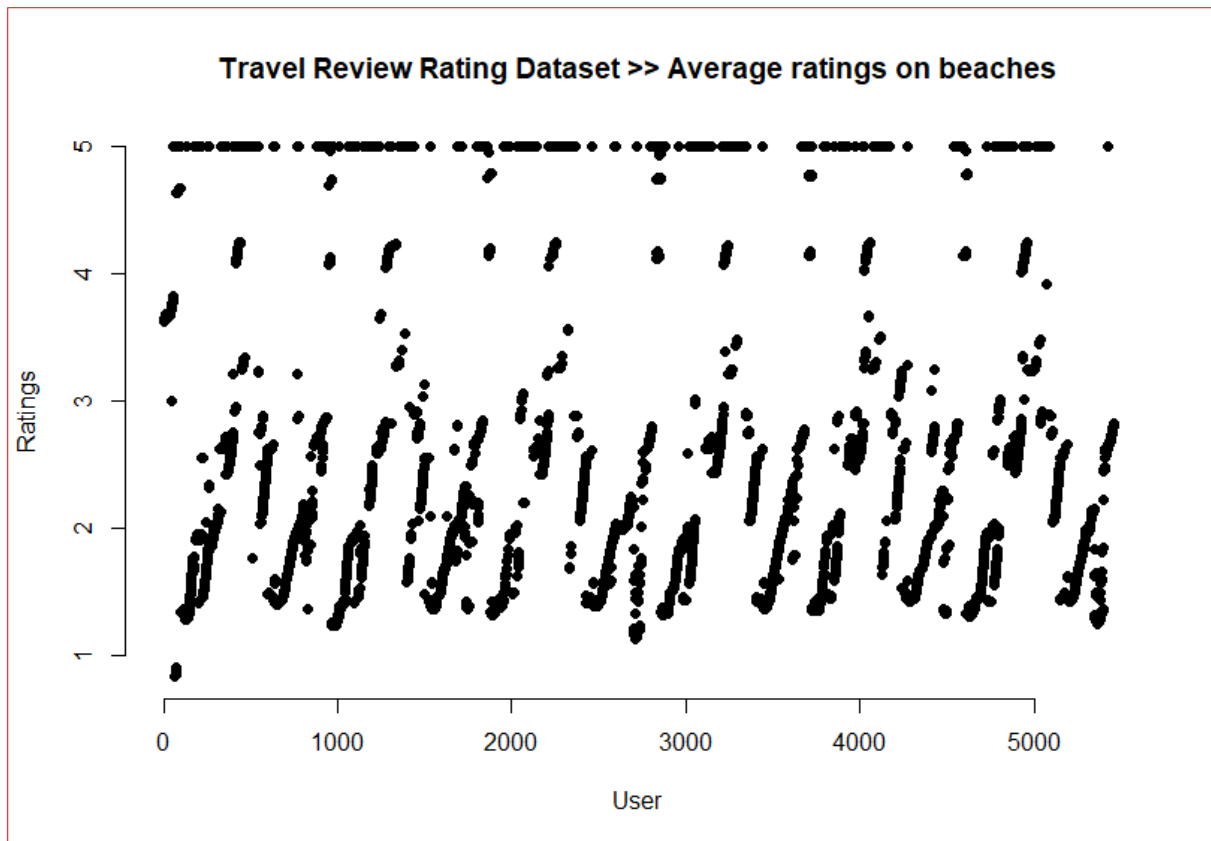
```
> hist(Google_Review_Ratings_.$'Average ratings on beaches', main = "Average ratings on  
beaches", xlab = "Review Ratings", ylab = "User" )
```



Boxplot แสดงแต่ละประเภทของการรีวิวของ User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน

Code สำหรับสร้าง Boxplot

```
> boxplot(Google_Review_Ratings_.$'Average ratings on beaches',
Google_Review_Ratings_.$'Average ratings on parks', Google_Review_Ratings_.$'Average ratings
on malls', Google_Review_Ratings_.$'Average ratings on beauty & spas',
Google_Review_Ratings_.$'Average ratings on cafes', names = c('Beaches','Park','Malls','Beauty &
Spas','Cafes'), xlab="Average ratings" )
```



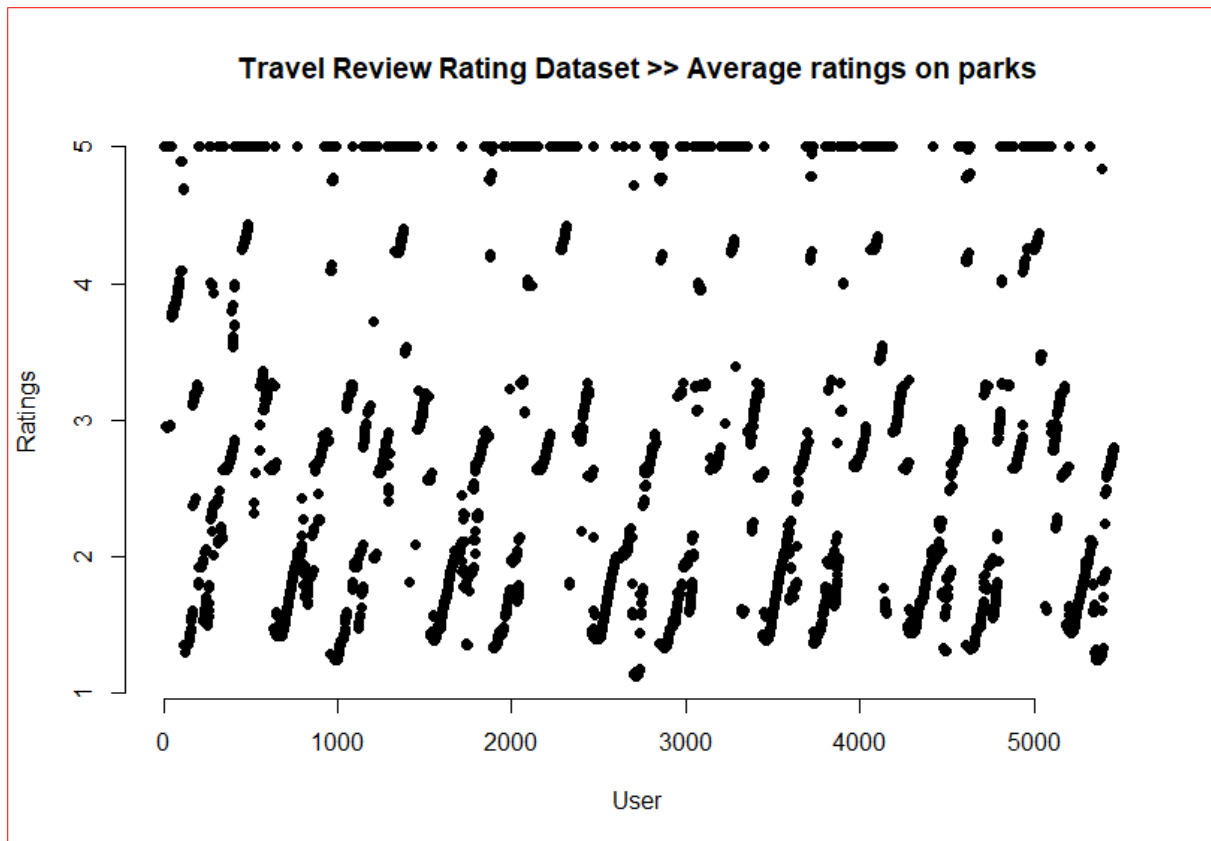
XY Scatter plot (Average ratings on beaches, User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)

Code สำหรับการ Plot XY Scatter Graph

```
> plot(Google_Review_Ratings_$'Average ratings on beaches',main="Travel Review Rating Dataset >> Average ratings on beaches",xlab = "user",ylab = "Ratings",pch=19, frame=FALSE)
> summary(Google_Review_Ratings_$'Average ratings on beaches')
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.830  1.730   2.460   2.797  4.093   5.000
```

ข้อมูลสถิติเบื้องต้นของ Average ratings on beaches จาก User ทั้งหมด พบว่า

- มี Min Ratings คือ 0.830
- มีค่า Mean คือ 2.797
- มี Max Ratings คือ 5.000
- มีค่าที่ตกใน Quartile ที่ 1 คือ 1.730
- มีค่าที่ตกใน Quartile ที่ 3 คือ 4.093



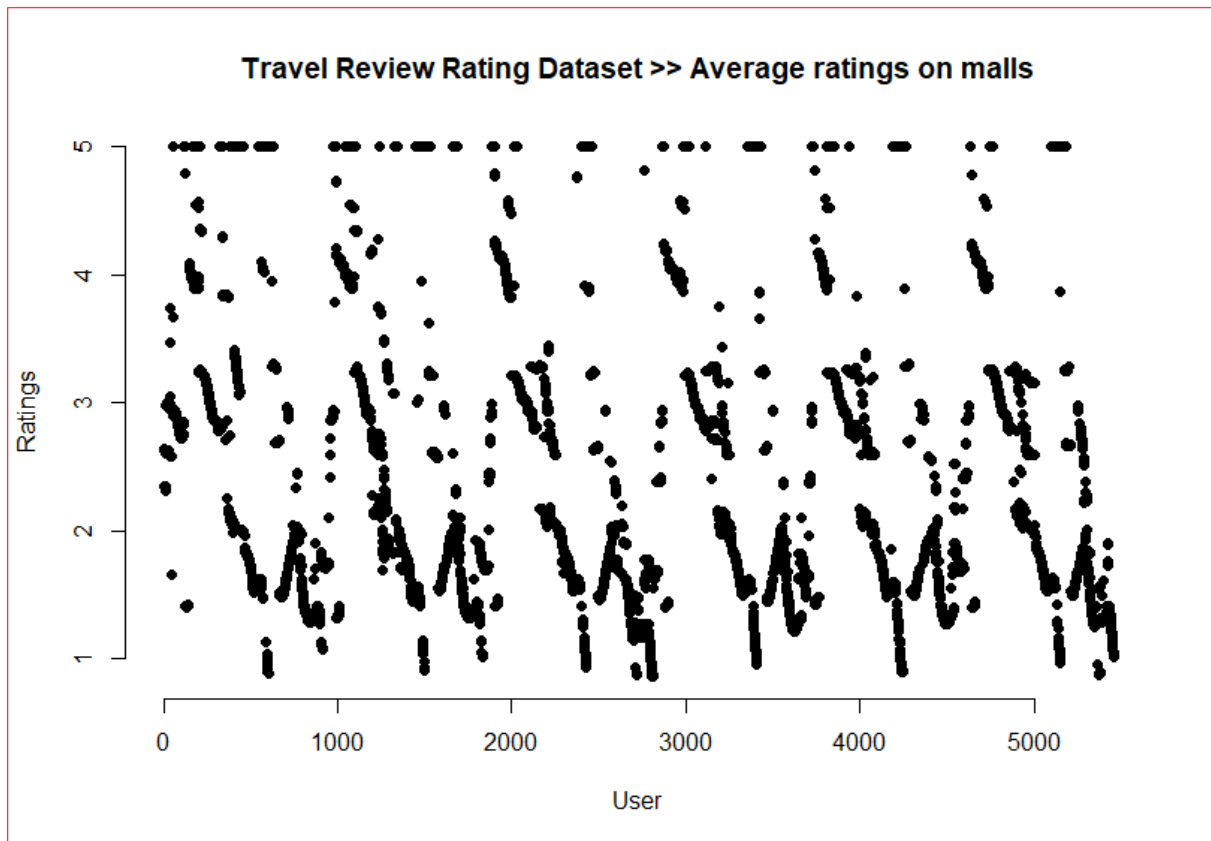
XY Scatter plot (Average ratings on parks, User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)

Code สำหรับการ Plot XY Scatter Graph

```
> plot(Google_Review_Ratings$`Average ratings on parks`,main="Travel Review Rating Dataset >> Average ratings on parks",xlab = "User",ylab = "Ratings",pch=19, frame=FALSE)
> summary(Google_Review_Ratings$`Average ratings on parks`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.120  1.770   2.670   2.959  4.312   5.000
```

ข้อมูลสถิติเบื้องต้นของ Average ratings on parks จาก User ทั้งหมด พบว่า

- มี Min Ratings คือ 1.120
- มีค่า Mean คือ 2.959
- มี Max Ratings คือ 5.000
- มีค่าที่ตกใน Quartile ที่ 1 คือ 1.770
- มีค่าที่ตกใน Quartile ที่ 3 คือ 4.312



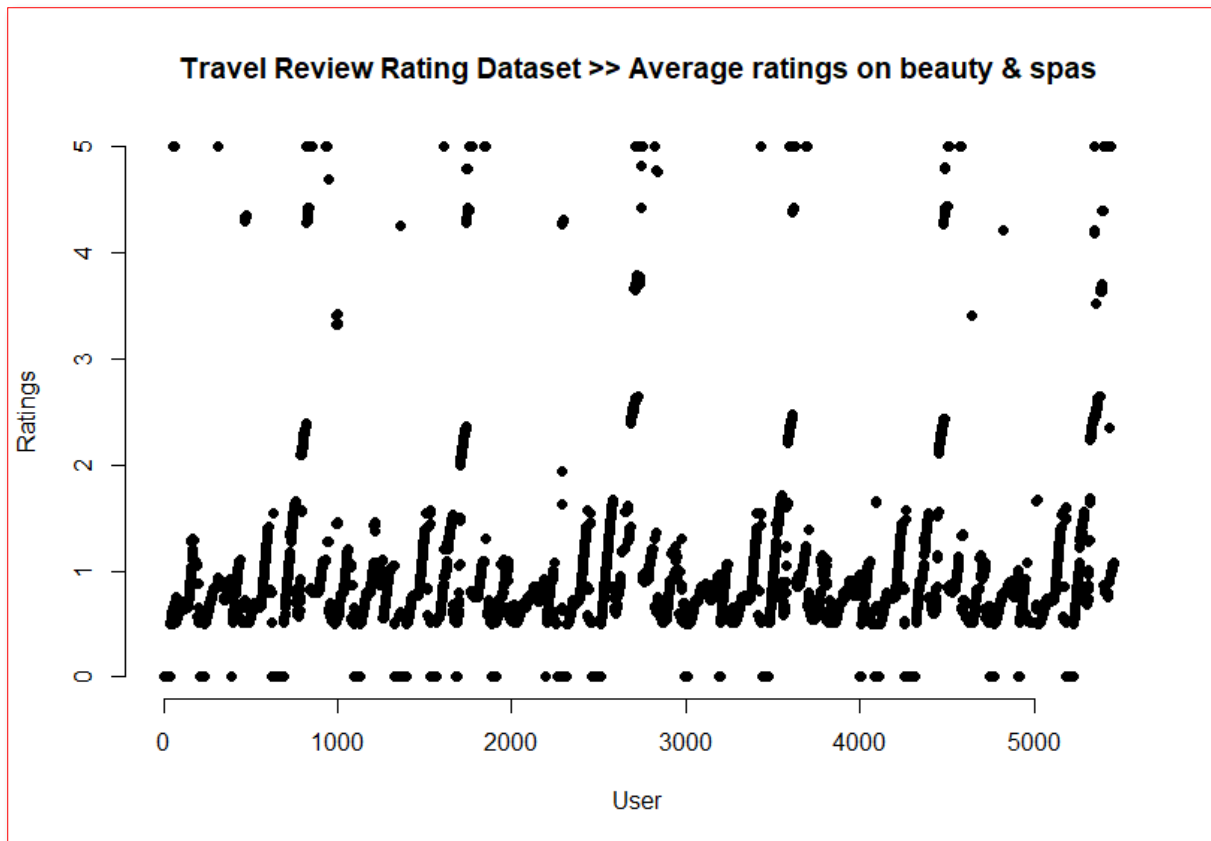
XY Scatter plot (Average ratings on malls, User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)

Code สำหรับการ Plot XY Scatter Graph

```
> plot(Google_Review_Ratings_$'Average ratings on malls',main="Travel Review Rating Dataset >> Average ratings on malls",xlab = "User",ylab = "Ratings",pch=19, frame=FALSE)
> summary(Google_Review_Ratings_$'Average ratings on malls')
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.860  1.620   2.170   2.541   3.190   5.000
```

ข้อมูลสถิติเบื้องต้นของ Average ratings on malls จาก User ทั้งหมด พบว่า

- มี Min Ratings คือ 0.860
- มีค่า Mean คือ 2.541
- มี Max Ratings คือ 5.000
- มีค่าที่ตกใน Quartile ที่ 1 คือ 1.620
- มีค่าที่ตกใน Quartile ที่ 3 คือ 3.190



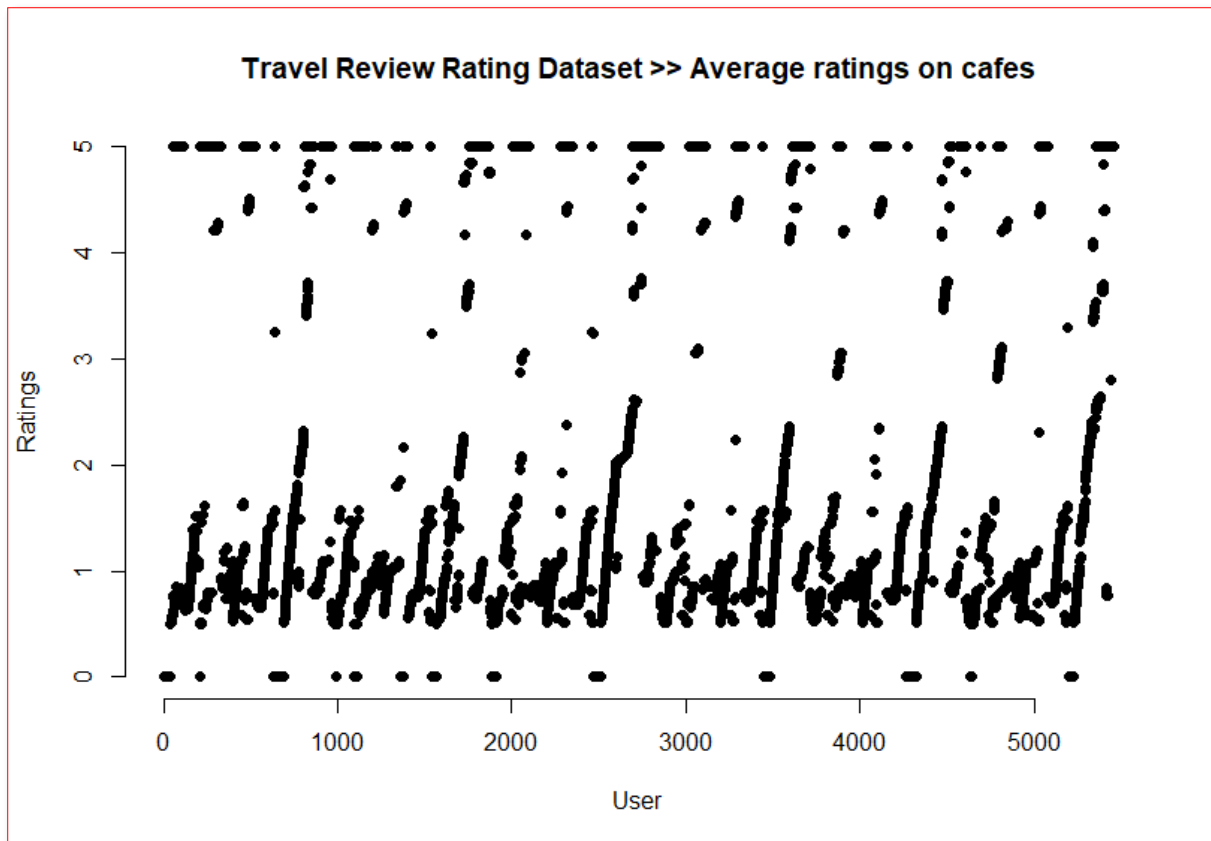
XY Scatter plot (Average ratings on beauty & spas, User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)

Code สำหรับการ Plot XY Scatter Graph

```
> plot(Google_Review_Ratings_.$`Average ratings on beauty & spas`,main="Travel Review Rating Dataset >> Average ratings on beauty & s
pas",xlab = "User",ylab = "Ratings",pch=19, frame=FALSE)
> summary(Google_Review_Ratings_.$`Average ratings on beauty & spas`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.5700  0.7600  0.9658  1.0000  5.0000
```

ข้อมูลสถิติเบื้องต้นของ Average ratings on beauty & spas จาก User ทั้งหมด พบว่า

- มี Min Ratings คือ 0.000
- มีค่า Mean คือ 0.9658
- มี Max Ratings คือ 5.000
- มีค่าที่ตกใน Quartile ที่ 1 คือ 0.570
- มีค่าที่ตกใน Quartile ที่ 3 คือ 1.000



XY Scatter plot (Average ratings on cafes, User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน)

Code สำหรับการ Plot XY Scatter Graph

```
> plot(Google_Review_Ratings$`Average ratings on cafes`,main="Travel Review Rating Dataset >> Average ratings on cafes",xlab = "User",ylab = "Ratings",pch=19, frame=FALSE)
> summary(Google_Review_Ratings$`Average ratings on cafes`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  0.740   1.030   1.751   2.070   5.000
```

ข้อมูลสถิติเบื้องต้นของ Average ratings on cafes จาก User ทั้งหมด พบว่า

- มี Min Ratings คือ 0.000
- มีค่า Mean คือ 1.751
- มี Max Ratings คือ 5.000
- มีค่าที่ตกใน Quartile ที่ 1 คือ 0.740
- มีค่าที่ตกใน Quartile ที่ 3 คือ 2.070

บทวิเคราะห์ข้อมูลจากกราฟ

ชุดข้อมูลนี้มาจากที่เก็บแมชชีนเลิร์นนิงของมหาวิทยาลัยแคลิฟอร์เนีย, เออร์ไวน์ (UC Irvine) : ข้อมูลการจัดอันดับรีวิวการเดินทาง ชุดข้อมูลนี้จะถูกเติมโดยการจับคะแนนของผู้ใช้จากรีวิวของ Google รีวิวเกี่ยวกับสถานที่ท่องเที่ยวจาก 24 หมวดหมู่ทั่วยุโรปได้รับการพิจารณา ทางผู้จัดทำได้หดยกคะแนนผู้ใช้ Google มีตั้งแต่ 1 ถึง 5 และมีการคำนวณคะแนนผู้ใช้เฉลี่ยต่อหมวดหมู่มาคิดวิเคราะห์

จากการวิเคราะห์กราฟ *XY Scatter plot* ทางผู้จัดทำจะแบ่งเป็น 5 ส่วนตามแต่ละประเภทของการรีวิว เพราะเนื่องจากเป็นข้อมูลที่ไม่มีความซับซ้อนมาก และนำจำนวน User ผู้รีวิวสถานที่เหล่านั้นจำนวน 5,456 คน มาเทียบกับข้อมูลสถิติเบื้องต้นของ Average ratings ทั้ง 5 ประเภท พบว่า User ส่วนมากที่เข้ามาเรีวนั้น มีการให้คะแนนเป็นไปในทางที่ดีเกี่ยวกับประเภท Ratings on parks เห็นได้จากคะแนนเฉลี่ยนั้นมีค่ามากที่สุดคือ 2.959 ส่วนรองลงมาจะเป็นประเภท Ratings on beaches ที่มีคะแนน 2.797 คะแนน และประเภท Ratings on malls ที่มีคะแนน 2.541 คะแนน แต่กลับกันจากกราฟยังพบอีกว่าค่าคะแนน Average ratings ของ Ratings on beauty & spas กับ Ratings on cafes ที่มีคะแนนการรีวิวเฉลี่ยต่ำมากคือ 0.965 คะแนน และ 1.751 คะแนน ตามลำดับ เนื่องจากความเป็นจริงสถานที่ที่ผู้คนส่วนใหญ่มักจะไปกันจะเป็นที่ที่เกี่ยวข้องกับธรรมชาติ อาจเป็นเพราะความลึกลับและความสวยงามของสถานที่เที่ยวทางธรรมชาติ ที่มีความน่าดึงดูดผู้คนมากกว่าสถานที่ที่ถูกสร้างขึ้นเองโดยมนุษย์ ซึ่งอาจมีข้อดีได้หลายปัจจัย อาทิ ได้รู้จักการวางแผน ได้ฝึกไหวพริบและรู้จักการแก้ปัญหาเฉพาะหน้า ได้ใกล้ชิดกับธรรมชาติและเข้าใจวิถีชีวิต ส่วนสถานที่เที่ยวรองลงมาคงหนีไม่พ้นกับ พวกประเภทห้างสรรพสินค้าต่าง ๆ เนื่องจากผู้คนที่มีความอยากได้สิ่งของที่จำเป็น อาทิ ยารักษาโรค และของอุปโภคบริโภค รวมถึงสินค้าแฟชั่นต่าง ๆ หรืออาจเป็นการผ่อนคลายทางด้านร่างกายจิตใจ และอารมณ์ เช่น การเข้าฟิตเนส การดูหนัง การเล่นเกม เป็นต้น และประเภทสุดท้ายจากกราฟคือ สถานที่ที่พักผ่อนหย่อนใจ อาทิ ร้านอาหารไทย ร้านสปา และคาเฟ่ต่าง ๆ เพราะมนุษย์นั้นมีความความรักสวย รักงามและมีเส้นทางที่จะเดินในแบบของตัวเอง