King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE213 Data Models, 2/2020

LAB Lecture 3: Basic types of data visualization

Assign Date: 5 Feb 2021 Due Date: 11 Feb 2021

Plot 3 graphs from the super store data and explain the meaning of each graph and what question required this graph as an answer

1. Download, import library and dataset.

```
[14] install.packages("hrbrthemes")

    Installing package into '/usr/local/lib/R/site-library'
    (as 'lib' is unspecified)

    also installing the dependencies 'extrafontdb', 'Rttf2pt1', 'extrafont'
```

```
library("tidyverse")
library("hrbrthemes")
```

NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
      Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
      if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow

```
[2] df_origin <- read.csv("http://fastdata.in.th/data-model-2021/superstore.csv", sep="|")
```

2. Preprocessing dataset and summary it.

```
[3] df_origin
```

21	CA-2011-143336	2012-08-27T00:00:00Z	2012-09-01T00:00:00Z	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	...	94
22	CA-2013-137330	2014-12-10T00:00:00Z	2014-12-14T00:00:00Z	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	...	680
23	CA-2013-137330	2014-12-10T00:00:00Z	2014-12-14T00:00:00Z	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	...	680
24	US-2014-156909	2015-07-17T00:00:00Z	2015-07-19T00:00:00Z	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	...	19
25	CA-2012-106320	2013-09-25T00:00:00Z	2013-09-30T00:00:00Z	Standard Class	EB-13870	Emily Burns	Consumer	United States	Orem	...	840
26	CA-2013-121755	2014-01-16T00:00:00Z	2014-01-20T00:00:00Z	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	...	900
27	CA-2013-121755	2014-01-16T00:00:00Z	2014-01-20T00:00:00Z	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	...	900

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE213 Data Models, 2/2020

LAB Lecture 3: Basic types of data visualization

Assign Date: 5 Feb 2021 Due Date: 11 Feb 2021

```
[4] df <- df_origin %>%
        mutate_if(is.character, as.factor)
```

```
[5] summary(df)
```

```
      Row.ID                Order.ID                      Order.Date
 Min.   :   1    CA-2014-100111:  14    2014-09-06T00:00:00Z:  38
 1st Qu.:2499    CA-2014-157987:  12    2015-09-03T00:00:00Z:  36
 Median :4998    CA-2013-165330:  11    2014-11-11T00:00:00Z:  35
 Mean   :4998    US-2013-108504:  11    2015-12-02T00:00:00Z:  34
 3rd Qu.:7496    CA-2012-131338:  10    2015-12-03T00:00:00Z:  34
 Max.   :9994    CA-2013-105732:  10    2015-12-10T00:00:00Z:  33
                 (Other)       :9926    (Other)             :9784
              Ship.Date               Ship.Mode      Customer.ID
 2013-12-16T00:00:00Z:  35    First Class   :1538    WB-21850:  37
 2015-09-27T00:00:00Z:  34    Same Day      : 543    JL-15835:  34
 2015-11-22T00:00:00Z:  32    Second Class  :1945    MA-17560:  34
 2015-12-07T00:00:00Z:  32    Standard Class:5968    PP-18955:  34
 2015-09-07T00:00:00Z:  30                           CK-12205:  32
 2015-09-16T00:00:00Z:  30                           EH-13765:  32
 (Other)             :9801                           (Other) :9791
              Customer.Name          Segment             Country
```

3. Plot a graph

   3.1. Heat map

   Code

```
[108] df %>%
        group_by(Region,Sub.Category) %>%
        summarise(total_quantity=sum(Quantity), .groups = 'drop') %>%
        ggplot(aes(Region, Sub.Category, fill=total_quantity)) +
        geom_tile() +
        scale_fill_gradient(name="Total quantity",low="black", high="red") +
        theme_ipsum() +
        ylab("Sub Category") +
        ggsave(file="heat_map_1.png", width=8, height=12, dpi=300)
```

King Mongkut's University of Technology Thonburi

Faculty of Engineering, Department of Computer Engineering

CPE213 Data Models, 2/2020

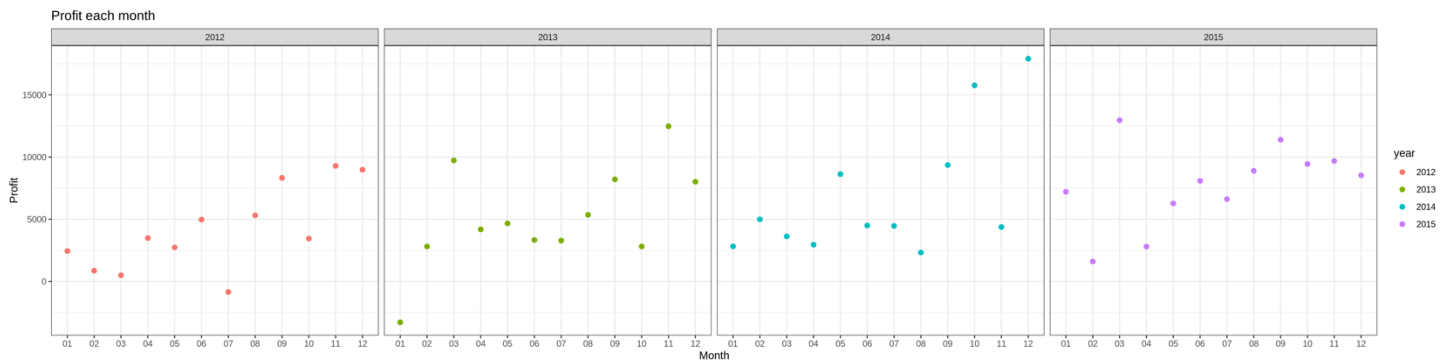LAB Lecture 3: Basic types of data visualization

Assign Date: 5 Feb 2021 Due Date: 11 Feb 2021

Graph
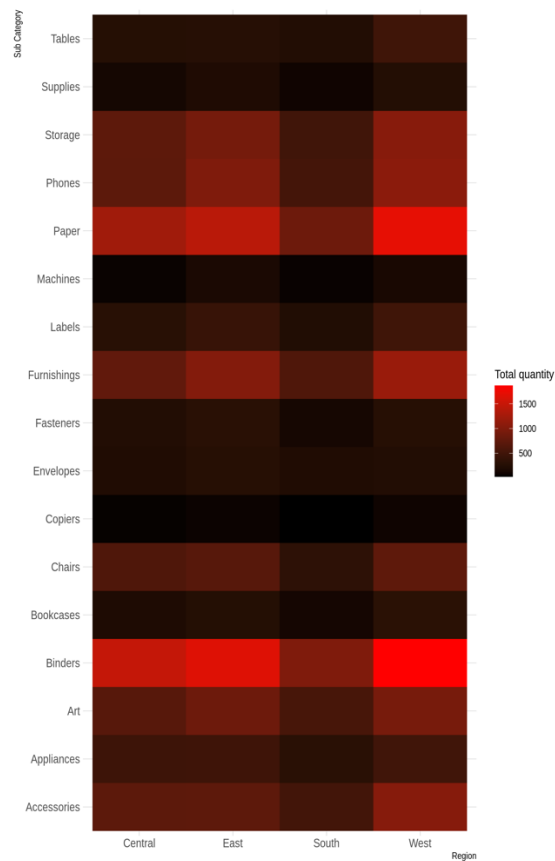
1) Meaning of this graph.

<u>Answer</u> แสดงความสัมพันธ์ของจำนวน ของที่ขายได้ในแต่ละหมวดหมู่ย่อยของแต่ละภูมิภาค

2) Question for this graph.

<u>Answer</u> สินค้าในแต่ละหมวดหมู่ย่อยที่ควรมีการกักตุ่นสินค้า ในแต่ละภูมิภาค เพื่อให้เพียงพอต่อการขาย?

3.2. Scatter plot

Code

```
df %>%
    mutate( year = strftime(as.Date(Order.Date, format="%Y-%m-%d"),"%Y"),
            month = strftime(as.Date(Order.Date, format="%Y-%m-%d"),"%m")) %>%
    group_by(year, month) %>%
    summarise(total_profit = sum(Profit), .groups = 'drop') %>%
    ggplot(aes(month, total_profit, color=year)) +
    geom_point(size=2) +
    facet_grid(.~year) +
    theme_bw() +
    ggtitle("Profit each month") +
    xlab("Month") +
    ylab("Profit") +
    ggsave(file="scatter_point_2.png", width=20, height=5, dpi=300)
```

Graph

1) Meaning of this graph.

<u>Answer</u> แสดงความสัมพันธ์ของกำไรในแต่ละเดือน

2) Question for this graph.

<u>Answer</u> เดือนไหนมียอดกำไรสูงที่สุดในแต่ละปี ? เพื่อจะได้ออกโปรโมชันเพื่อดึงดูดความสนใจลูกค้า

3.3. Bar plot

Code

```
df %>%
    group_by(Segment,Sub.Category) %>%
    summarise(total_sales = sum(Sales), .groups='drop') %>%
    ggplot(aes(Sub.Category, total_sales, fill = Segment)) +
    geom_col() +
    theme_bw() +
    ggtitle("Total sales each sub category compare to each segment") +
    xlab("Sub Category") +
    ylab("Total Sales") +
    ggsave(file="bar_plot_3.png", width=15, height=5, dpi=300)
```

Graph
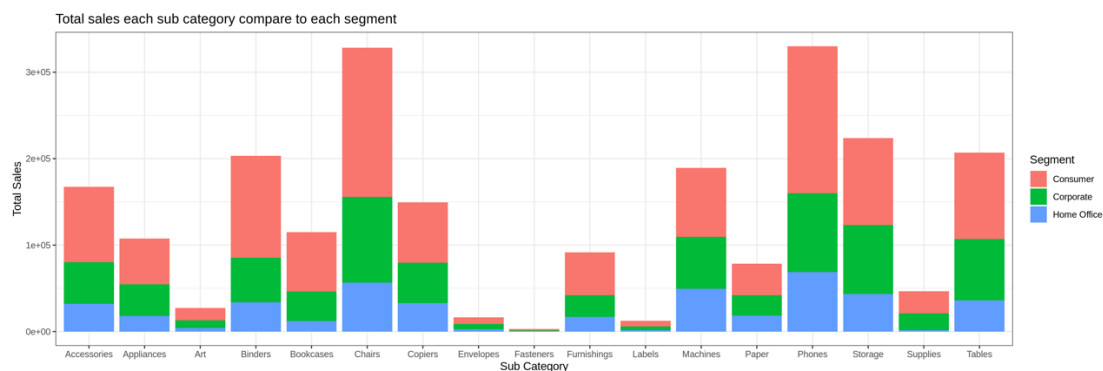
1) Meaning of this graph.

   <u>Answer</u> แสดงความสัมพันธ์ของยอดขายในแต่ละหมวดหมู่ย่อยของสินค้า โดยเทียบกันในแต่ละ Segment

2) Question for this graph.

   <u>Answer</u> ในแต่ละหมวดหมู่ย่อยของสินค้าอันไหนมียอดขายน้อยที่สุดในแต่ละ Segment? เพื่อจะได้ออกนโยบายกระตุ้นการขาย