Prepared by Heejae Roh
Presented to Professor Behzad Ahmadi

Oct 28th, 2022

Nike vs Adidas

Northeastern University

Northeastern University

01

Nike vs Adidas

It's always interesting story. In this analysis,
I will analyze dataset of rating, review, pricing etc. about Nike vs Adidas

02

Rating & Price

Comparing rating, price data between
Adidas 3 branch vs Nike & Adidas vs Nike

03

New Data & Correlation

Create new data under condition
To make more precise and deeper analysis

04

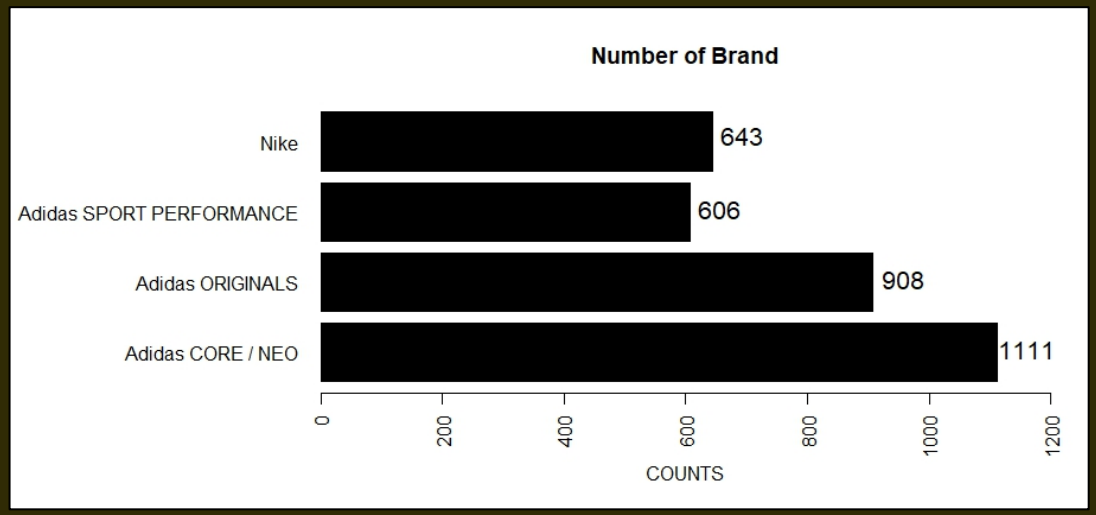
Learning & Question

Assume consumer loyalty between two
Finding pricing strategies and advice for two

Nike vs Adidas

Part1
Summarize Data

'Nike vs Adidas.csv' from Kaggle
3268 obs, of **10** variables



Nike(643) vs Adidas(2625)
Consumer reaction & price info about products

Explaining product variable(10) like
- Brand: 3 Adidas branch, 1 Nike
- Rating: evaluation of product
- Review: number of written
- Listing Price: original price
- Sale Price: discounted price

...

Excel & R data

	A	B	C	D	E	F	G	H	I	J	K
1	Product Na	Product ID	Listing Pric	Sale Price	Discount	Brand	Descriptio	Rating	Reviews	Last Visited	
2	Women's a	AH2430	14999	7499	50	Adidas OR	Channeling	4.8	41	2020-04-13T15:06:14	
3	Women's a	G27341	7599	3799	50	Adidas OR	A modern	3.3	24	2020-04-13T15:06:15	
4	Women's a	CM0081	999	599	40	Adidas CO	These adid	2.6	37	2020-04-13T15:06:15	
5	Women's a	B44832	6999	3499	50	Adidas CO	Inspired by	4.1	35	2020-04-13T15:06:15	

```
> str(nkadds)                                          ***restructuring like R
'data.frame':          3268 obs. of  10 variables:
$Product.Name, $Product.ID, $Listing.Price, $Sales.Price, $Discount, $Brand, $Description, $Rating,
$Reviews, $Last.Visited
```

Part1

Rating

What is mean and median of rating

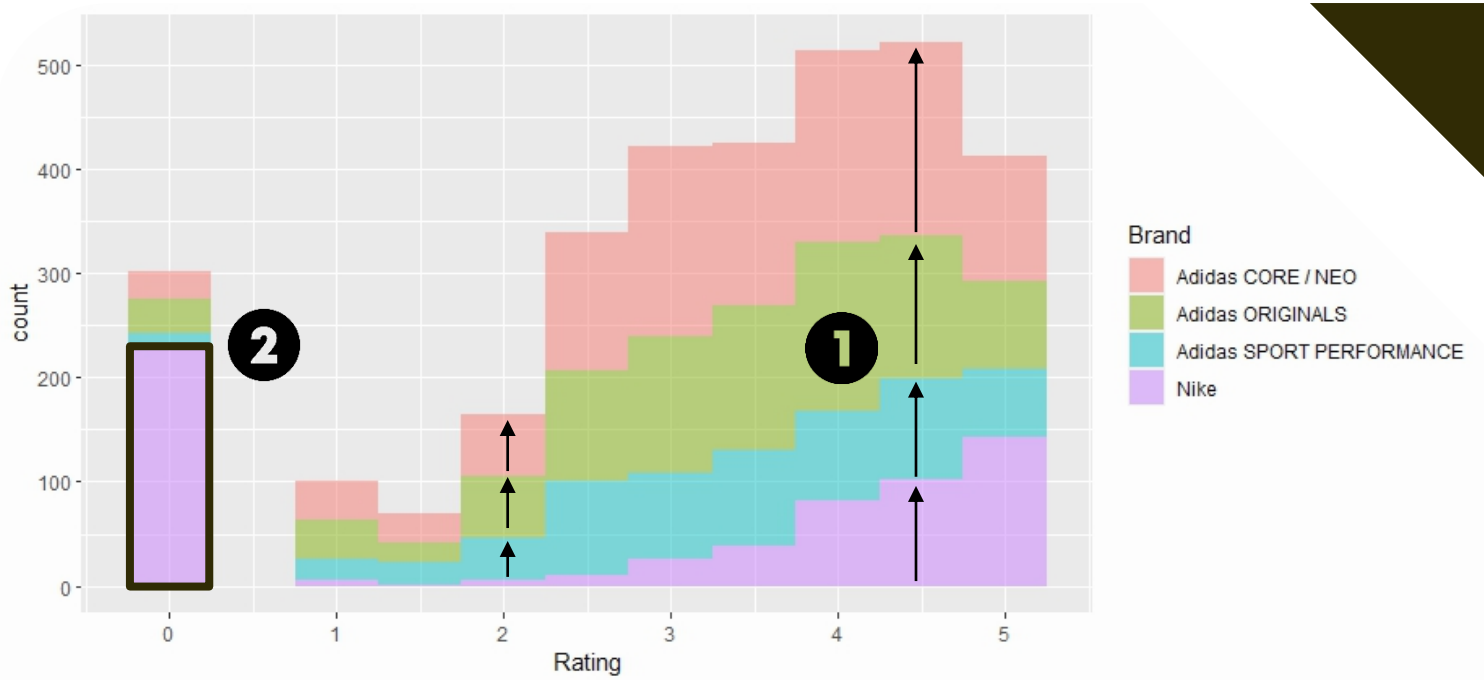
Brand	Mean	Median
Adidas CORE / NEO*	3.41	3.5
Adidas ORIGINALS	3.32	3.5
Adidas SPORT PERFORMANCE	3.35	3.4
Nike(all)	2.73	3.8

3

*Cleaning Data
Change wrong name ex) Adidas Adidas -> Adidas

Brief Discussion

1. Rating counts high in higher points
2. Nike rating get many 0 portions
   -> might be many not viewed Obs.
3. [Median] Nike > 3 Adidas branch
   [Mean] Nike < 3 Adidas branch



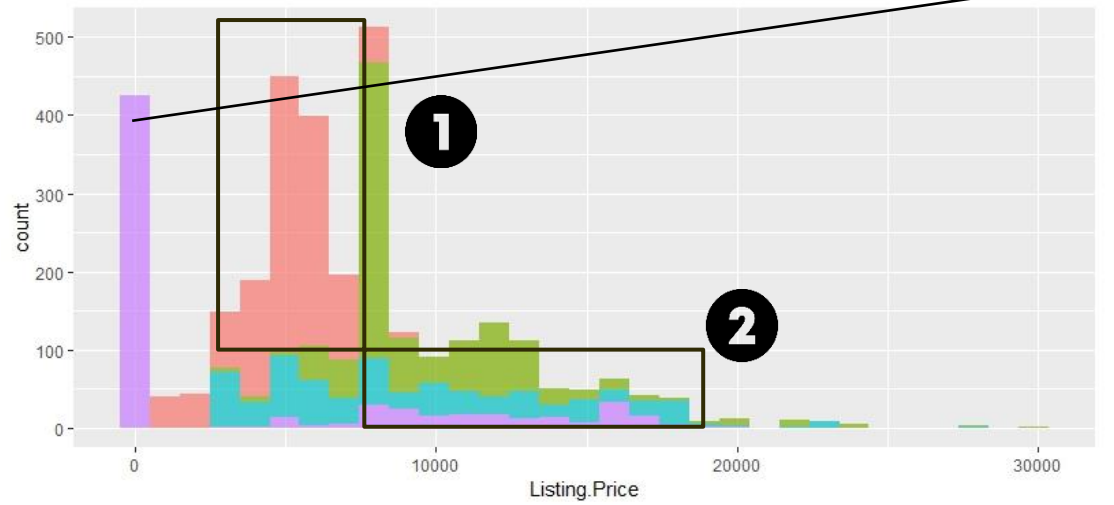
Part1

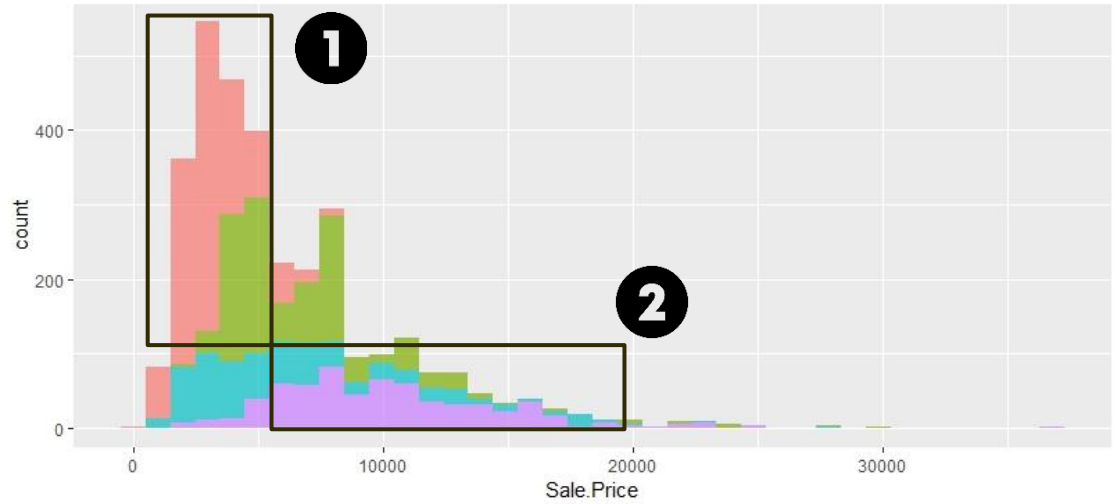
Price

<Price & Discount rate>

Brand	List	Sale	rate
Adidas CORE / NEO	4952	3116	37%
Adidas ORIGINALS	9820	6812	31%
Adidas SPORT PERFORMANCE	9132	6322	31%
Nike**	11484	10213	11%

**Cleaning Data
There are '0's in 'Listing.Price' of Nike
-> calculate the average mean without 0 price
No Discount rate for Nike
-> calculate in R with making fuction





Brief Discussion

1. A. CORE / NEO is lowest price and highest Discount rate
2. Price of Nike is highest and Discount rate is lowest

Part2

New 'Adidas' vs Nike**

Brand	Mean Rating
Nike(all)	2.7
Nike**	4.1
Adidas	3.4

1.4↑

❶

Brand	Median Rating
Nike(all)	3.8
Nike**	4.4
Adidas	3.5

0.6↑ Apply 0 portion (2.7549%) of Adidas

❷

Create Data

Using code which change every adidas branch name as 'Adidas'

- Make new column 'Brand.name' which contains only 'Adidas' & 'Nike'

** calculate the average 0 portion of Adidas

0 portion of Adidas product= 2.7549% of total

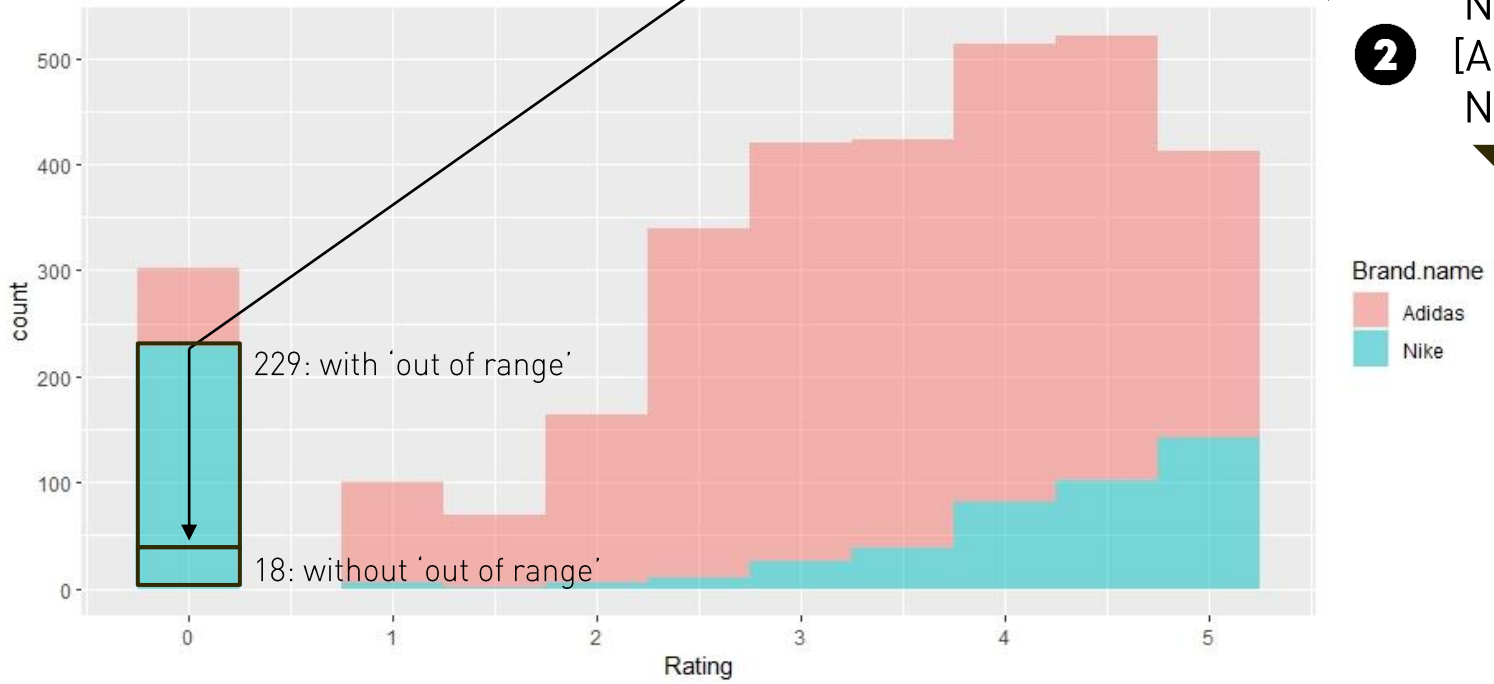
-> apply 0 portion of Adidas to Nike Mean & Median

-> 18 of 643 = 2.7% of total and make other NA

Brief Discussion

❶ [Adjusted mean]
Nike(4.1) > Adidas(3.4)

❷ [Adjusted median]
Nike[4.4] > Adidas[3.8]



Part2

New Rating

Using review as criteria

Cleaning Data

change rating 0 to NA, under condition 'Review is less than 10'

code: data$Rating[Rating == 0 & Reviews < 10] <- NA

- Make new data: without [rating= 0 & review < 10]
- Becuase 0 rating and less review might means it could be visited not enoughly by consumers

There are 236 values which is "out of range"
: 0 rating and less than 10 review
mostly in Nike

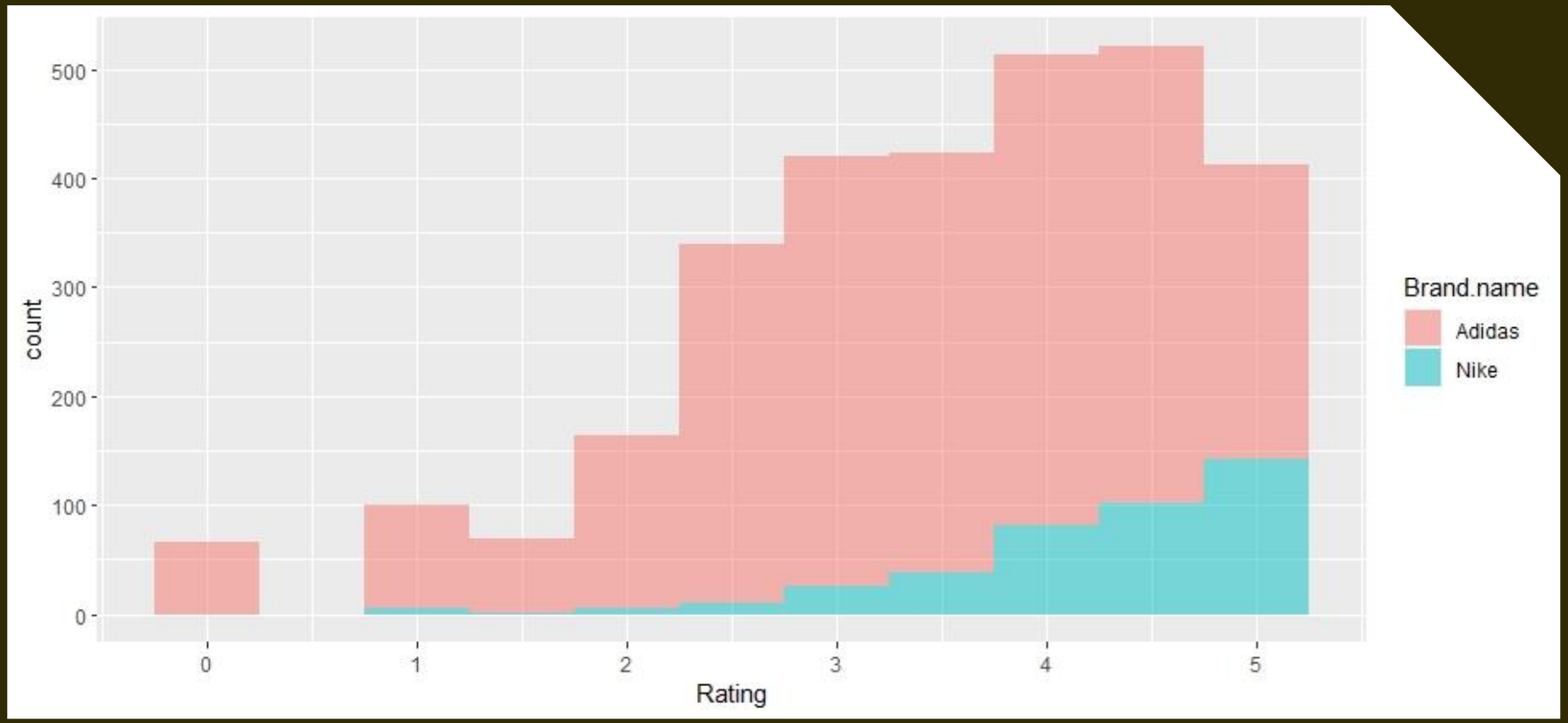
<Mean>

Brand	Rating
Nike	4.3
Adidas	3.4

<Median>

Brand	Rating
Nike	4.4*
Adidas	3.5

*same with original median



Part2 & 3

Correation

Cleaning data: in original dataset '0' for all Nike Discount column
-> Calculate Discount in R with Listing.Price and Sale.Price

Analysis adidas

1. Cor(Sale.Price & Discount) = -0.52
   → Huge, but logically obvious
2. Cor(another & the other) = close to zero
   → Rating & S.P is not in lenear Cor.

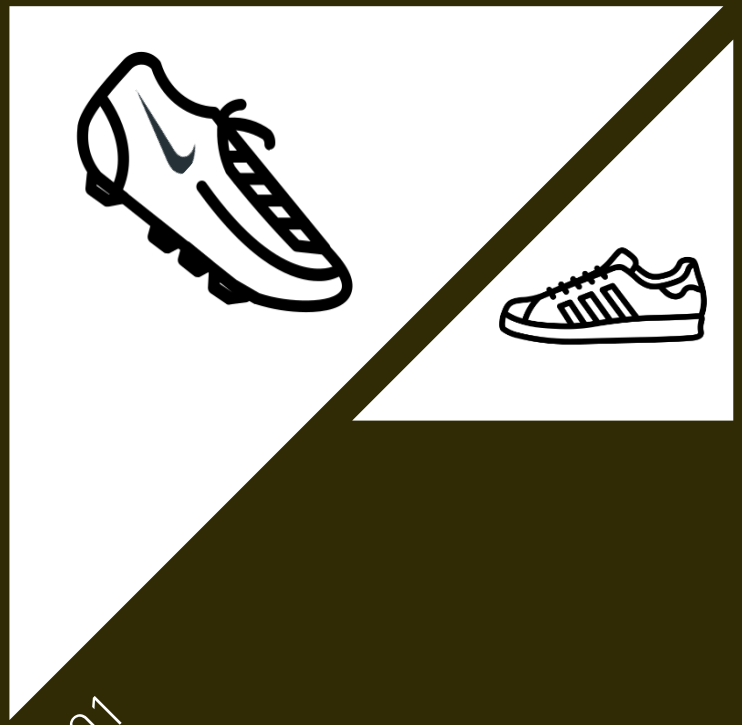
Analysis NIKE

3. Cor(Rating & Review) = 0.3
   → 'Rating high' ⇄ 'Review number high'
4. Cor(Discount & Sales Price) = 0.88
   → Ignore this value, cuz lot of 0 Listing price

**Listing price exist = discount exist -> high correlation

**Unusual Value: Sale.Price and Listing.Price relation is obvious

Learning & Follow-up Question



LEARN 01

In this dataset made sites,
'Adidas' consumer make rating more actively.
-> Nike needs to change ways of exposure and reaction

LEARN 02

Mean of 'Adidas' rating is higher than Nike's. Omitting '0' ratings,
Nike is higher than Adidas
-> Reacting consumers' rate is higher in Nike, even price of Nike is higher

Question 01

Q. Which consumer's loyalty is higher?
Nike loyalty might be higher, cuz higher rating portion is high in LEARN2
-> Correlation Review & Rating(0.3) also explains this(page 7)

Question 02

Q. What is the pricing strategy of Nike?
Lower discount rate. High listing price, and high sale price in this dataset

Question 03

Q. What is the pricing strategy of Adidas?
In mean price, Adidas is in cheap pricing, and high discount rate.

Advice

- **Nike needs to check exposure**
- **Adidas needs to manage rating**
- **Both need to reexamine pricing strategy**

Bibliography

- KASH. (2020). Adidas vs nike. kaggle. Retrieved from https://www.kaggle.com/datasets/kaushiksuresh147/adidas-vs-nike

- Cookbook for R. (n.d.). Plotting distributions (ggplot2). Retrieved from http://www.cookbook-r.com/Graphs/Plotting_distributions_(ggplot2)/

- STHDA. (n.d.). ggplot2 : Quick correlation matrix heatmap - R software and data visualization. Retrieved from http://www.sthda.com/english/wiki/ggplot2-quick-correlation-matrix-heatmap-r-software-and-data-visualization