Assignment 4 – MBI 664

Contents

[Question 1 3](#_Toc25329969)

[R Code: 3](#_Toc25329970)

[Answer: 3](#_Toc25329971)

[Question 2 3](#_Toc25329972)

[R Code: 3](#_Toc25329973)

[Answer: 3](#_Toc25329974)

[Question 3 3](#_Toc25329975)

[R Code: 3](#_Toc25329976)

[Answer: 3](#_Toc25329977)

[Question 4 4](#_Toc25329978)

[R Code: 4](#_Toc25329979)

[Answer: 4](#_Toc25329980)

[Question 5 4](#_Toc25329981)

[Answer: 4](#_Toc25329982)

[Question 6 5](#_Toc25329983)

[R Code: 5](#_Toc25329984)

[Answer: 5](#_Toc25329985)

[Question 7 6](#_Toc25329986)

[R Code: 6](#_Toc25329987)

[Answer: 6](#_Toc25329988)

[Question 8 7](#_Toc25329989)

[Answer: 7](#_Toc25329990)

[Question 9 8](#_Toc25329991)

[R Code: 8](#_Toc25329992)

[Answer: 8](#_Toc25329993)

[Comments: 8](#_Toc25329994)

[R Code: 9](#_Toc25329995)

[Answer: 9](#_Toc25329996)

[Comments: 9](#_Toc25329997)

[Question 10 10](#_Toc25329998)

[R Code: 10](#_Toc25329999)

[Answer: 10](#_Toc25330000)

[Question 11 11](#_Toc25330001)

[R Code: 11](#_Toc25330002)

[Answer: 11](#_Toc25330003)

[Question 12 12](#_Toc25330004)

[R Code: 12](#_Toc25330005)

[Answer: 12](#_Toc25330006)

[Question 13 13](#_Toc25330007)

[R Code: 13](#_Toc25330008)

[Answer: 13](#_Toc25330009)

[Question 14 14](#_Toc25330010)

[R Code: 14](#_Toc25330011)

[Answer: 14](#_Toc25330012)

[R Code: 15](#_Toc25330013)

[Answer: 15](#_Toc25330014)

[R Code: 16](#_Toc25330015)

[Answer: 16](#_Toc25330016)

[Question 15 17](#_Toc25330017)

[R Code: 17](#_Toc25330018)

[Answer: 17](#_Toc25330019)

# Question 1

Read the satisfaction.csv file into R. What does head=TRUE mean while reading the CSV file? (R code and text answer)

## R Code:

setwd("/Users/selva/Desktop/MBI/Fall 2019/MBI 664 Data Visuallization/W12")

library("ggplot2")

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

## Answer:

Head =TRUE specifies first row as column headers.

# Question 2

Display the first few lines of the dataset. (R code + Output)

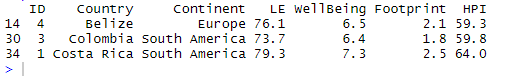
## R Code:

w <- read.table(file="Satisfaction.csv", head=TRUE, sep=",")

retval <- subset(w,ID<11)

print(retval)

## Answer:



# Question 3

Display the column names of the dataset. (R code + Output)

## R Code:

w <- read.table(file="Satisfaction.csv", head=TRUE, sep=",")

colnames(w)

## Answer:



# Question 4

Write the R code necessary to produce the following plot using the ggplot2 library. Note the Title in the graph for this question and all other question requiring a graph your title should include your name and the question that you are answering. (R code + graph)

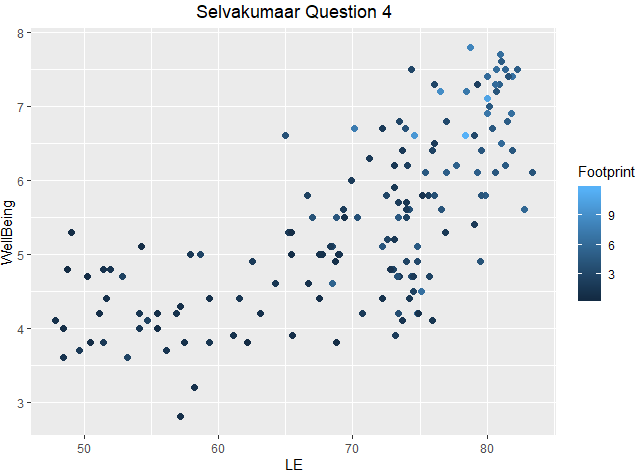
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x=LE,y=WellBeing,color=Footprint))

p + geom\_point(size=2) + ggtitle("Selvakumaar Question 4") + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 5

Write a paragraph (at least four sentences) on your interpretation of the above plot. Use the correlation analysis discussion from Stephen Few, if you think it is appropriate (Text)

## Answer:

The graph above shows a weak positive correlation between the two variables LE & WellBeing. This is because the data points are loosely grouped around the line of regression (line of best fit). It can be clearly seen that there is an increasing trend in the relationship of two variables however the values are scattered. Also the scatter plot reveals that the relationship can be a positive exponential correlation since the data points seems to follow an exponential growth curve as we move from one value to other.

# Question 6

Jitter the above plot by adding a geom layer to your ggplot2 object. (R code + graph)

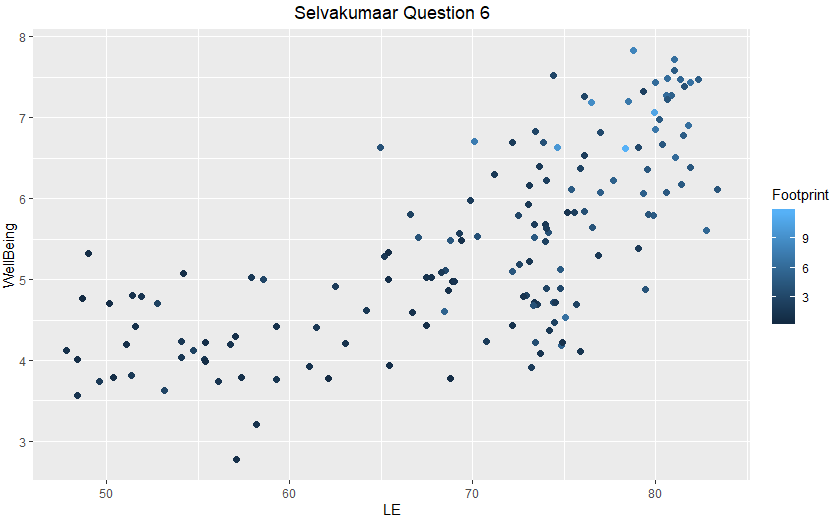
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x=LE,y=WellBeing,color=Footprint))

p + geom\_jitter(size=2) + ggtitle("Selvakumaar Question 6") + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 7

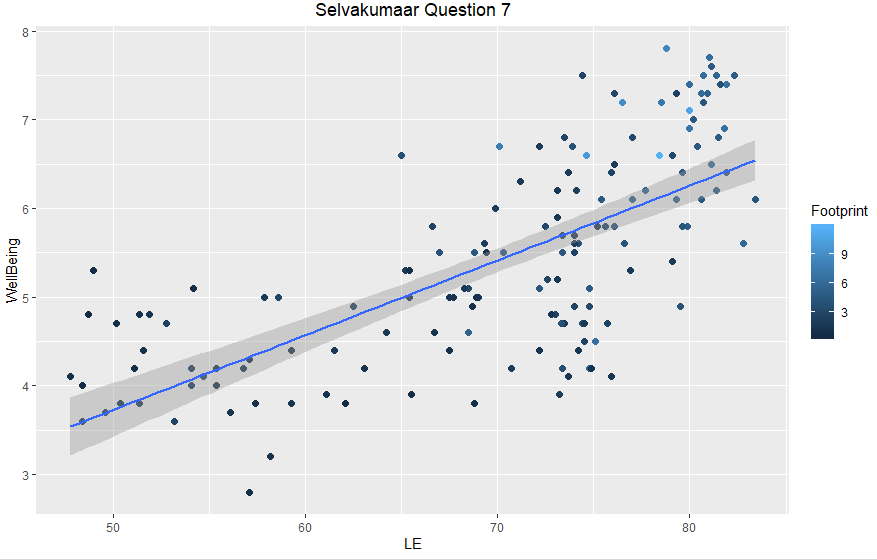
Now write the code necessary using the ggplot2 library to produce the following plot. (R code + graph)

## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x=LE,y=WellBeing,color=Footprint))

p + geom\_point(size=2) + geom\_smooth(method = "lm", se = TRUE) + ggtitle("Selvakumaar Question 7") + theme(plot.title = element\_text(hjust = .5))

Answer:

# Question 8

What does the line represent in the above graph? What does a residual mean with reference to the above graph? What does the band around the line indicate? (Text)

## Answer:

The line in the graph above represents Regression Line (also called as Smooth Line). It is obtained by using the method “lm” in geom\_smooth() function. “lm” refers to Linear Regression Model – this means that in the graph above Smoothing Method used was Linear Regression Model.

Residual is the deviation of each observed value from the mean (Regression Line). The band around the line represents the confidence interval. This was obtained by specifying se=TRUE in the geom\_amooth() function. By default R uses 95% as the Level of Confidence for a prediction in a Linear Model.

# Question 9

Explain the difference between aesthetic mapping and parameter setting – first using an example by writing R code and then write couple sentences to explain the difference (Code and two graphs)

Aesthetic Mapping:

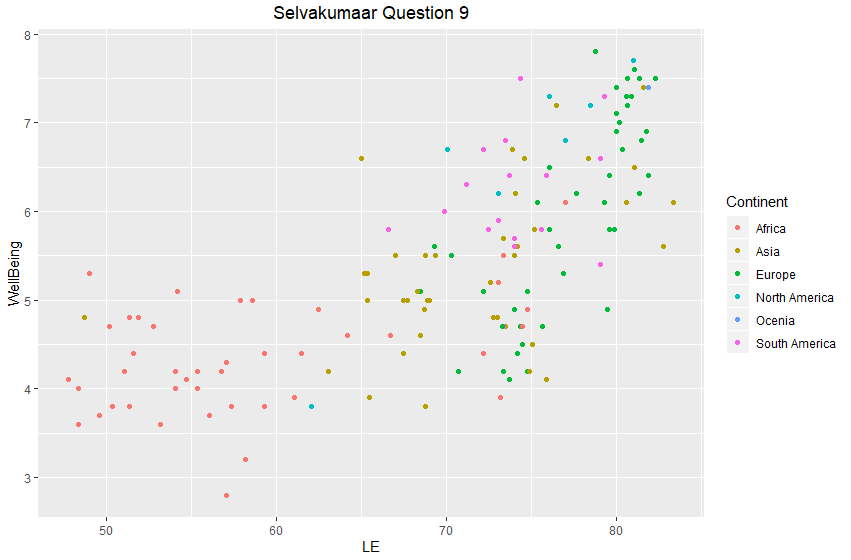
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x=LE,y=WellBeing))

p + geom\_point(aes(color=Continent)) + ggtitle("Selvakumaar Question 9") + theme(plot.title = element\_text(hjust = .5))

## Answer:



## Comments:

In Aesthetic mapping function (aes), the value of data in column (continent) decides the color of visual graphic (points in graph). As you can see from the above example, the points are displayed in different colors depending on the values in “Continent” column.

Parameter setting:

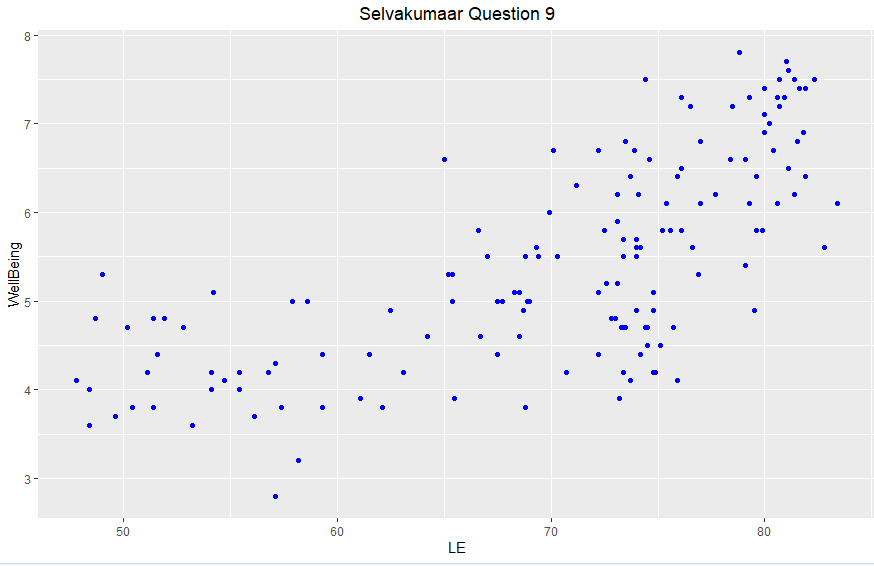
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x=LE,y=WellBeing))

p + geom\_point(color="Blue") + ggtitle("Selvakumaar Question 9") + theme(plot.title = element\_text(hjust = .5))

## Answer:



## Comments:

A constant value determines the visual representation of each data point. As you can see from the above example each data point is colored in blue (a constant color code) irrespective of column.

# Question 10

Produce the following plot again using the ggplot2 library. (R code and graph)

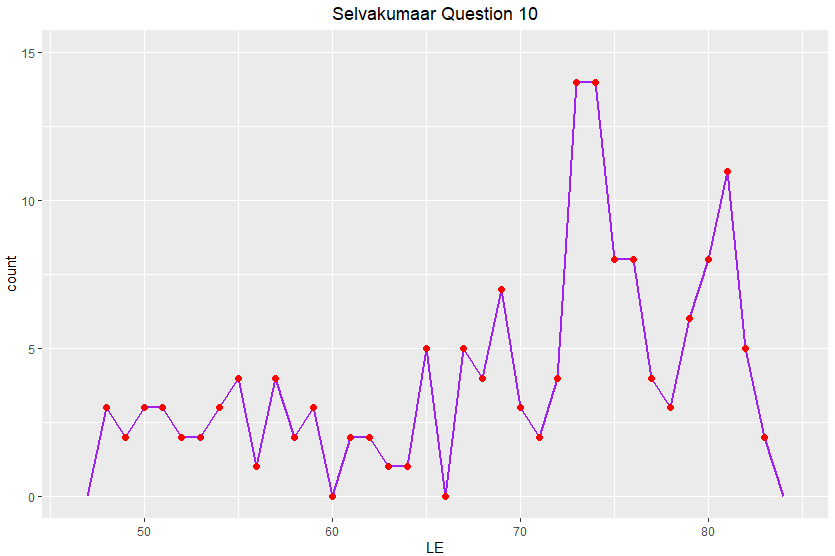
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x=LE))

p + geom\_freqpoly(color="purple",size=1,binwidth=1) + geom\_point(stat = "bin",color="red",size=2,binwidth=1) + ylim(0,15) + ggtitle("Selvakumaar Question 10") + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 11

Produce the following graph noting that the y-axis is now the proportion of the count of countries in each continent (from our dataset) and the y-axis scale ranges from 0.0 to 0.3, and the fill color is now red. (R code + graph)

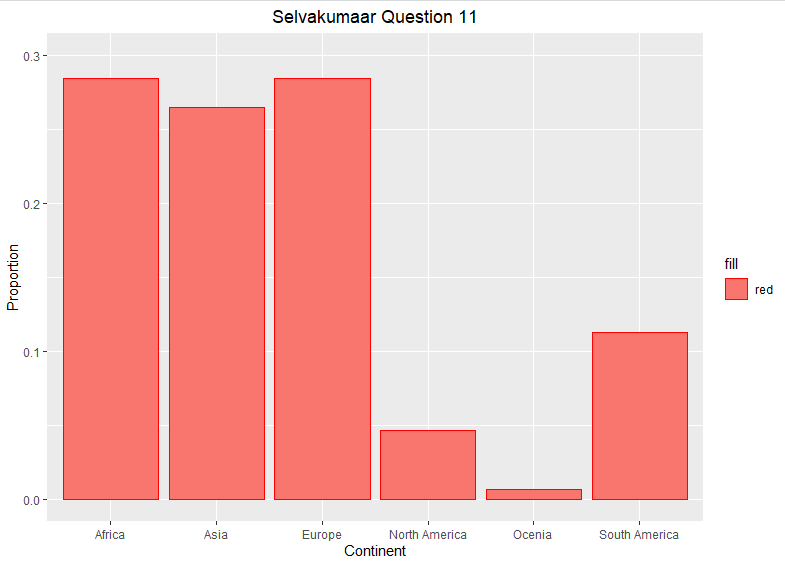
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x= Continent, fill = "red"))

p + geom\_bar (aes(y=(..count../sum(..count..))), color="red") + labs(title="Selvakumaar Question 11", x="Continent", y="Proportion") + ylim(0.0,0.3) + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 12

Use parameter mapping to produce the following graph where each continent is colored differently. (R code + graph)

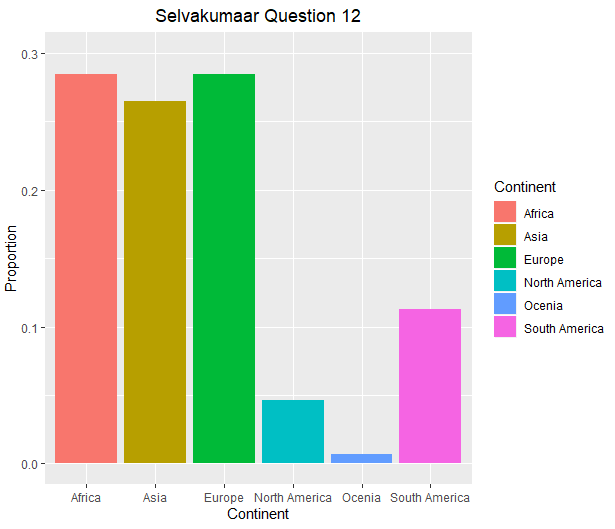
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x= Continent, fill = Continent))

p + geom\_bar (aes(y=(..count../sum(..count..)))) + labs(title="Selvakumaar Question 11", x="Continent", y="Proportion") + ylim(0.0,0.3) + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 13

Now, change the y-axis back to count and color intensity of the bars to reflect the count and produce the following graph. (R code + graph)

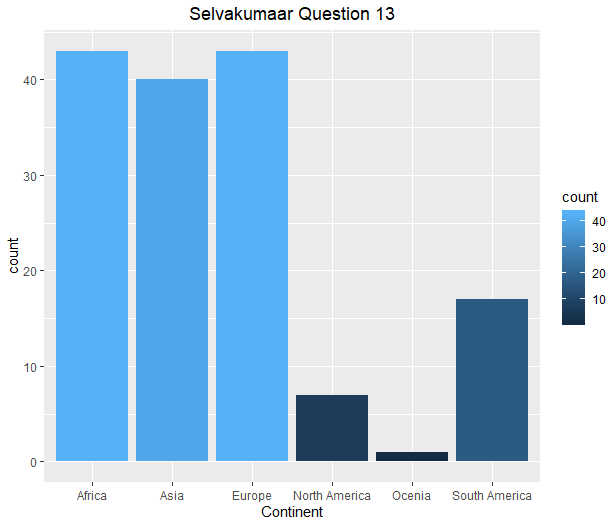
## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

p <- ggplot(data=w,aes(x= Continent))

p + geom\_bar (aes(fill = ..count..)) + labs(title="Selvakumaar Question 13") + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 14

Define a new variable for the WellBeing index for values greater than 5. Now, plot a stacked bar graph with Wellbeing >5 and <5 stacked, these bars next to each other, a filled graph that shows the proportions of these variables for each continent as shown below (R code and Graph):

Stacked Bars one above the other

## R Code:

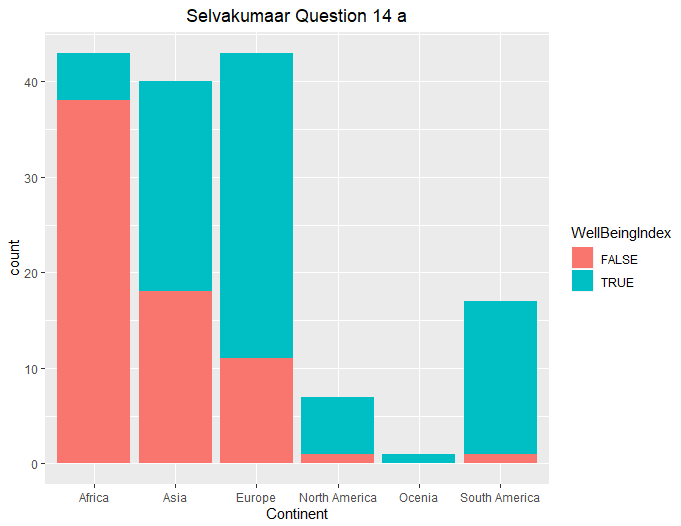
w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

w$WellBeingIndex <- w$WellBeing>5

p <- ggplot(data=w,aes(x= Continent))

p + geom\_bar (position = position\_stack(reverse = TRUE), aes(fill = WellBeingIndex)) + labs(title="Selvakumaar Question 14 a") + theme(plot.title = element\_text(hjust = .5))

## Answer:



Stacked bar chart side by side

## R Code:

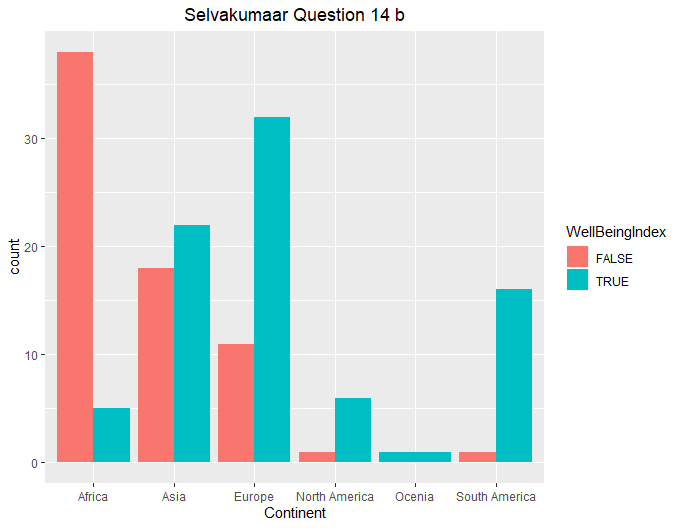
w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

w$WellBeingIndex <- w$WellBeing>5

p <- ggplot(data=w,aes(x= Continent))

p + geom\_bar (position = position\_dodge(), aes(fill = WellBeingIndex)) + labs(title="Selvakumaar Question 14 b") + theme(plot.title = element\_text(hjust = .5))

## Answer:



Filled Bar Chart:

## R Code:

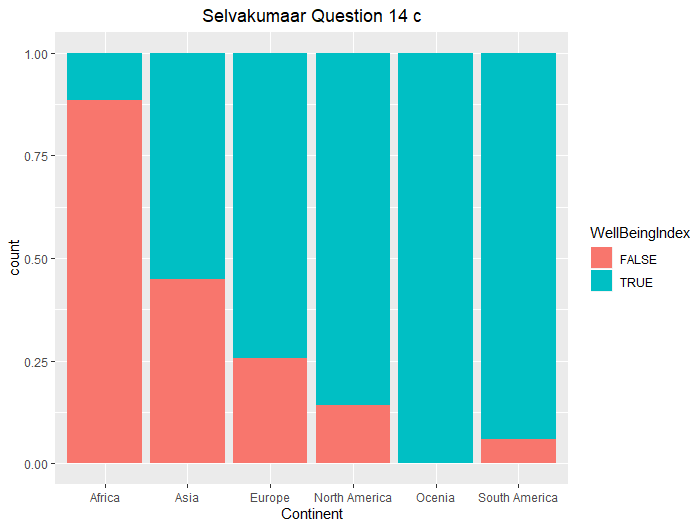
w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

w$WellBeingIndex <- w$WellBeing>5

p <- ggplot(data=w,aes(x= Continent))

p + geom\_bar (position = position\_fill(reverse = TRUE), aes(fill = WellBeingIndex)) + labs(title="Selvakumaar Question 14 c") + theme(plot.title = element\_text(hjust = .5))

## Answer:



# Question 15

Produce the first graph shown above with NKU colors instead of as shown above (Black and Gold) R code and graphic object.

## R Code:

w <- read.csv(file="Satisfaction.csv", head=TRUE, sep=",")

w$WellBeingIndex <- w$WellBeing>5

p <- ggplot(data=w,aes(x= Continent))

p + geom\_bar (position = position\_stack(reverse = TRUE), aes(fill = WellBeingIndex)) + scale\_fill\_manual(values=c("Black","Gold")) + labs(title="Selvakumaar Question 15") + theme(plot.title = element\_text(hjust = .5))

## Answer:

