## 64036\_Assignment\_1

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

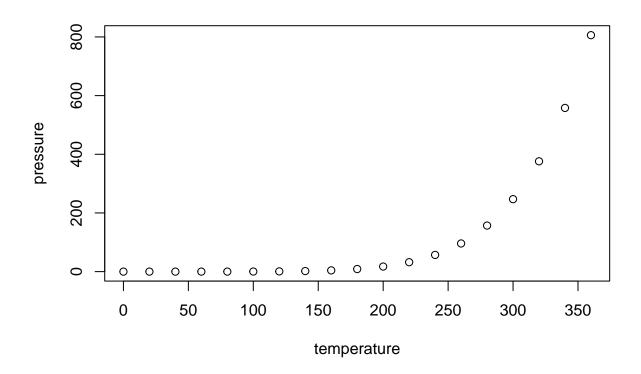
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

## summary(cars)

```
##
                         dist
        speed
           : 4.0
                              2.00
##
    Min.
                    Min.
                           :
    1st Qu.:12.0
                    1st Qu.: 26.00
##
    Median:15.0
                    Median: 36.00
##
    Mean
            :15.4
                    Mean
                           : 42.98
                    3rd Qu.: 56.00
##
    3rd Qu.:19.0
    Max.
            :25.0
                           :120.00
                    Max.
```

## **Including Plots**

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

summary(Online\_Retail)

head(Online Retail) summary(Online Retail)

X1\_test <- Online\_Retail\$Country summary(X1\_test) head(Online\_Retail) spec(Online\_Retail) summary(Online\_Retail)

colSums(is.na(Online\_Retail))

table(grepl("^C", Online Retail\$InvoiceNo))

 $table (Online\_RetailCountry) revenue < -Online_RetailCountry) \%>\% \\ summarise \\ (sales=sum(Online\_RetailQuantity*Online\_RetailUnitPrice))$ 

revenue <- Online\_Retail %>% group\_by(Online\_Retail\$Country) %>% summarise (Country\_Count = n()) %>% mutate(country\_percentage = (Country\_Count/sum(Country\_Count))\*100) %>% filter(country\_percentage>1)

 $str(revenue) \ str(Online\_Retail) \ revenue \ \%>\% \ select(sales) \ precoftotal <- \ revenue \ sales * sum(revenue sales)$ 

 $\label{eq:contraction} \begin{array}{lll} \text{head}(\text{precoftotal}) & \text{Percents} & <& \text{Online\_Retail} & \% > \% & \text{c}(\text{Online\_Retail} & \text{Country}, Online_RetailQuantity}, \\ \text{Online\_Retail} & UnitPrice) & \text{head}(Percents) & \text{summary}(Percents) & & -c(Online_RetailCountry}, \\ \text{Online\_Retail} & & \text{Quantity}, Online_RetailUnitPrice}) & \text{summary}(Percent\_1) \\ \end{array}$ 

 $percent\_2 <- Online\_Retail \%>\% \ select (Online\_Retail Country, Online_Retail Quantity, Online\_Retail \$UnitPrice)$ 

 ${\tt revenue}["percents"] < - {\tt revenue} sales/sum(sales) \ {\tt summary}(revenue) \ {\tt sum}(revenue \$ sales)$ 

summary(Online\_Retail)

revenue <- Online\_Retail %>% group\_by(Online\_RetailCountry)table(Online\_RetailCountry)

Online Retail <- Online Retail %>% mutate(TransactionValue = Quantity \* UnitPrice)

Online\_Retail %>% group\_by(Country) %>% summarise (sum\_TransactionValue = sum(TransactionValue)) %>% filter (sum\_TransactionValue > 130000)

Online\_Retail %>% filter(Country == "Germany")

Germany <- Online\_Retail %>% filter(Country == "Germany") hist(Germany\$TransactionValue,main = paste("Histogram of Germany Transactions"))

plot(Germany)

 $\label{lis.na} On line\_Retail $$ CustomerID), ] \%>\% group\_by (CustomerID) \%>\% summarise (Customer\_Count = n(), sum\_total= sum(TransactionValue)) \%>\% arrange (desc(Customer\_Count))$ 

Online\_Retail [!is.na(Online\_Retail\$CustomerID),] %>% group\_by(CustomerID) %>% summarise (Customer\_Count = n(),sum\_total= sum(TransactionValue)) %>% arrange (desc(sum\_total))

colMeans(is.na(Online Retail)\*100)

 $\label{lem:contine} Online\_Retail \%>\% \ group\_by(Country) \%>\% \ summarise(CustomerID\_Missing = sum(is.na(CustomerID))) \%>\% \ filter(CustomerID\_Missing>0)$ 

Online Retail Cancelled <- Online Retail %>% filter(Country=="France",Quantity<0) %>% count()

 $Online\_Retail\_Total <- Online\_Retail \%>\% \ filter(Country == "France") \%>\% \ count()$ 

 $(Online\_Retail\_Cancelled n/Online_Retail\_rotaln)*100$ 

(149/8557)\*100

 $\label{lem:condine_Retail} Online\_Retail \%>\% \ group\_by(StockCode) \%>\% \ summarise(sum\_transactionvalue = sum(TransactionValue)) \\ \%>\% \ arrange(desc(sum\_transactionvalue))$ 

length(unique(Online Retail\$CustomerID))

Temp=strptime(Online Retail\$InvoiceDate,format='\%m/\%d/\%Y \%H:\%M',tz='GMT')

Online Retail\$New Invoice Date <- as.Date(Temp)

Online\_Retail $Invoice_Day_Week = weekdays(Online_RetailNew_Invoice_Date)$ 

Online\_Retail\$New\_Invoice\_Hour = as.numeric(format(Temp, "%H"))

Online Retail\$New Invoice Month = as.numeric(format(Temp, "%m"))

Online\_Retail %>% group\_by(Invoice\_Day\_Week) %>% summarise(number\_of\_transactions=n()) %>% mutate(trans\_percent = (number\_of\_transactions/sum(number\_of\_transactions))\*100)