Ch 2 : Credit ggplot

Kiran Khanderao

April 25, 2019

# knitr::opts\_chunk$set(echo = TRUE)  
  
## R Markdown  
#This is an R Markdown document.  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

library(readxl)

## Warning: package 'readxl' was built under R version 3.5.3

#install.packages("xlsx")  
#getwd()  
##.......Chapter 2.....Credit Data ggplot........##  
rm(credit\_ds)

## Warning in rm(credit\_ds): object 'credit\_ds' not found

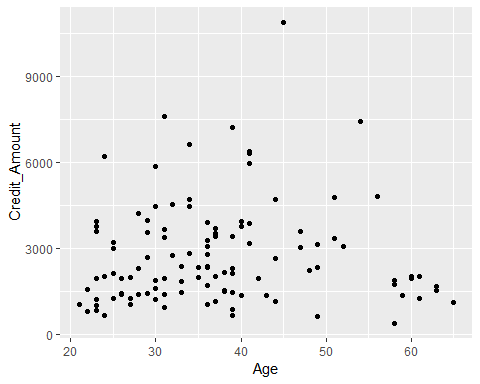
credit\_ds <- read\_excel("credit.xlsx")  
summary(credit\_ds)

## Creditability Account\_Balance Duration\_Credit\_Months Payment\_Status   
## Min. :0.000 Min. :1.000 Min. : 6.00 Min. :0.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:12.00 1st Qu.:2.000   
## Median :1.000 Median :2.000 Median :18.00 Median :4.000   
## Mean :0.955 Mean :2.649 Mean :19.61 Mean :3.117   
## 3rd Qu.:1.000 3rd Qu.:4.000 3rd Qu.:24.00 3rd Qu.:4.000   
## Max. :1.000 Max. :4.000 Max. :48.00 Max. :4.000   
## Purpose Credit\_Amount Value\_Savings\_Stocks  
## Min. : 0.000 Min. : 385 Min. :1.000   
## 1st Qu.: 0.000 1st Qu.: 1448 1st Qu.:1.000   
## Median : 2.000 Median : 2241 Median :1.000   
## Mean : 2.486 Mean : 2763 Mean :2.198   
## 3rd Qu.: 3.000 3rd Qu.: 3576 3rd Qu.:4.000   
## Max. :10.000 Max. :10875 Max. :5.000   
## Length\_of\_Current\_Employment Instalment\_pct Gender\_Marital\_Status  
## Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :3.000 Median :4.000 Median :3.000   
## Mean :3.541 Mean :2.937 Mean :2.748   
## 3rd Qu.:5.000 3rd Qu.:4.000 3rd Qu.:3.000   
## Max. :5.000 Max. :4.000 Max. :4.000   
## Guarantors Duration\_at\_Current\_Address Most\_Valuable\_Asset  
## Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000   
## Median :1.000 Median :3.000 Median :2.000   
## Mean :1.081 Mean :2.829 Mean :2.099   
## 3rd Qu.:1.000 3rd Qu.:4.000 3rd Qu.:3.000   
## Max. :3.000 Max. :4.000 Max. :4.000   
## Age Concurrent\_Credits Type\_of\_Apartment  
## Min. :21.00 Min. :1.000 Min. :1.000   
## 1st Qu.:29.00 1st Qu.:3.000 1st Qu.:2.000   
## Median :36.00 Median :3.000 Median :2.000   
## Mean :37.19 Mean :2.892 Mean :1.865   
## 3rd Qu.:41.50 3rd Qu.:3.000 3rd Qu.:2.000   
## Max. :65.00 Max. :3.000 Max. :3.000   
## No\_of\_Credits\_To\_Bank Occupation Dependents Telephone   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:3.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :1.000 Median :3.000 Median :1.000 Median :1.000   
## Mean :1.459 Mean :2.838 Mean :1.153 Mean :1.459   
## 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:2.000   
## Max. :4.000 Max. :4.000 Max. :2.000 Max. :2.000   
## Foreign\_Worker   
## Min. :1.000   
## 1st Qu.:1.000   
## Median :1.000   
## Mean :1.054   
## 3rd Qu.:1.000   
## Max. :2.000

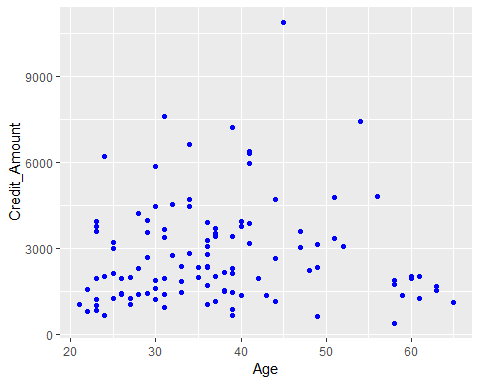
str(credit\_ds)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 111 obs. of 21 variables:  
## $ Creditability : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Account\_Balance : num 1 1 2 1 1 1 1 1 4 2 ...  
## $ Duration\_Credit\_Months : num 18 9 12 12 12 10 8 6 18 24 ...  
## $ Payment\_Status : num 4 4 2 4 4 4 4 4 4 2 ...  
## $ Purpose : num 2 0 9 0 0 0 0 0 3 3 ...  
## $ Credit\_Amount : num 1049 2799 841 2122 2171 ...  
## $ Value\_Savings\_Stocks : num 1 1 2 1 1 1 1 1 1 3 ...  
## $ Length\_of\_Current\_Employment: num 2 3 4 3 3 2 4 2 1 1 ...  
## $ Instalment\_pct : num 4 2 2 3 4 1 1 2 4 1 ...  
## $ Gender\_Marital\_Status : num 2 3 2 3 3 3 3 3 2 2 ...  
## $ Guarantors : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Duration\_at\_Current\_Address : num 4 2 4 2 4 3 4 4 4 4 ...  
## $ Most\_Valuable\_Asset : num 2 1 1 1 2 1 1 1 3 4 ...  
## $ Age : num 21 36 23 39 38 48 39 40 65 23 ...  
## $ Concurrent\_Credits : num 3 3 3 3 1 3 3 3 3 3 ...  
## $ Type\_of\_Apartment : num 1 1 1 1 2 1 2 2 2 1 ...  
## $ No\_of\_Credits\_To\_Bank : num 1 2 1 2 2 2 2 1 2 1 ...  
## $ Occupation : num 3 3 2 2 2 2 2 2 1 1 ...  
## $ Dependents : num 1 2 1 2 1 2 1 2 1 1 ...  
## $ Telephone : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Foreign\_Worker : num 1 1 1 2 2 2 2 2 1 1 ...

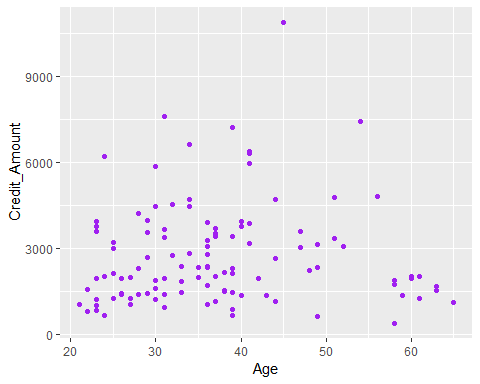
## GGplot  
ggplot(credit\_ds, aes(x = Age, y = Credit\_Amount)) +geom\_point()



ggplot(credit\_ds, aes(x = Age, y = Credit\_Amount)) +geom\_point(color="blue")



ggplot(credit\_ds, aes(Age,Credit\_Amount))+geom\_point(color="purple")



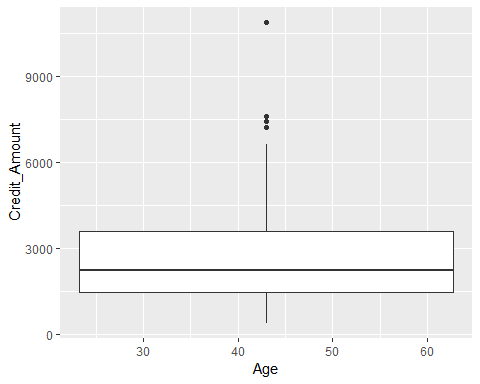
# something wrong in the plot ....  
  
ggplot(credit\_ds, aes(Age)) +geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

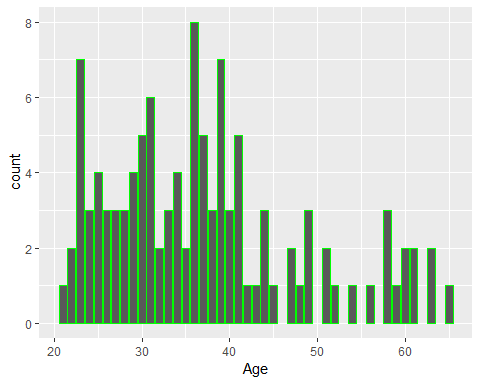


ggplot(credit\_ds, aes(Age,Credit\_Amount)) + geom\_boxplot()

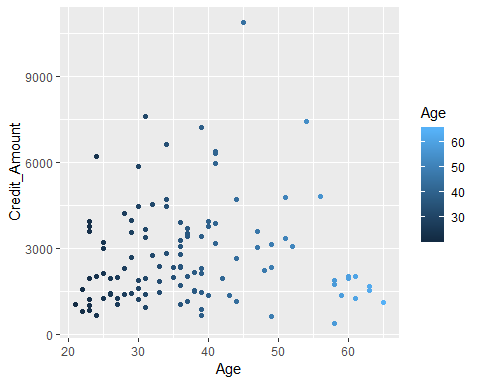
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?



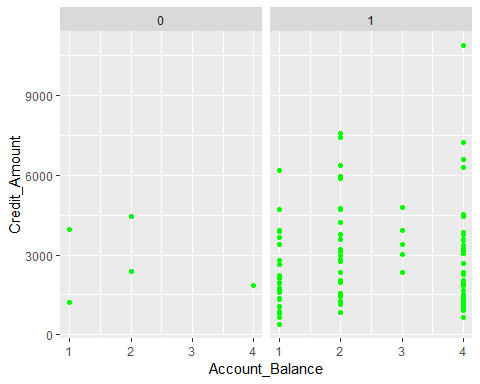
# Heart\_Attack\_Diagnosis = Age  
# Serum\_Credit\_Amountesterol = Credit\_Amount   
  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
ggplot(credit\_ds,aes(Age))+geom\_bar(colour="Green ")



# Additional aesthetical componenets like color,shape can be added.   
# Color added to depict 1: Absence 2: Presence of heart risk   
ggplot(credit\_ds, aes(Age,Credit\_Amount,colour=Age))+geom\_point()

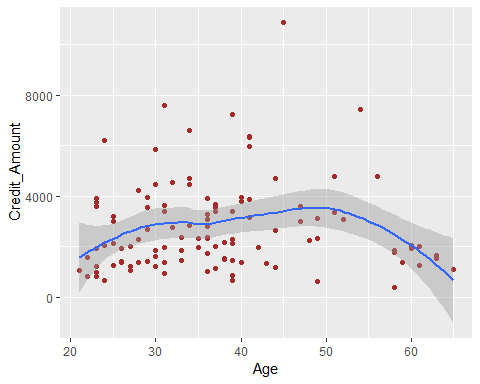


#Facetting used to replicate the display for the two groups:  
# 1 Absence of Risk 2 Presence of Risk   
ggplot(credit\_ds,aes(Account\_Balance,Credit\_Amount))+geom\_point(color="green")+facet\_wrap(~Creditability)



# Adding another layer,a smoother over the Age-Credit\_Amountesterol plot   
ggplot(credit\_ds,aes(Age, Credit\_Amount))+geom\_point(colour="brown")+geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



#`geom\_smooth()` using method = 'loess' and formula 'y ~ x'  
  
  
# Span parameter helps overlay a fitted curve on the scatterplot using the de fualt loess method and then linear model.  
# Possibly overfitted.   
ggplot(credit\_ds,aes(Age,Credit\_Amount))+geom\_point(colour="red")+geom\_smooth(span=0.15)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at 37

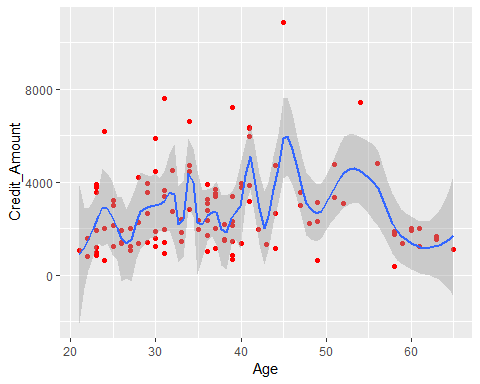
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 0

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used  
## at 37

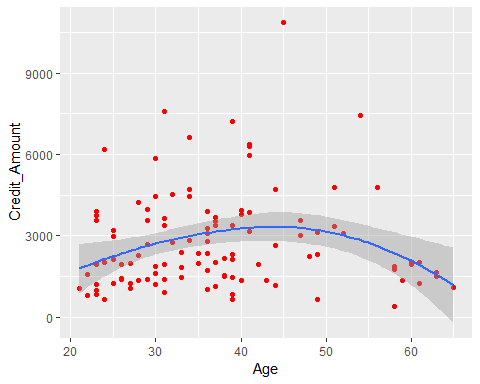
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 1

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal  
## condition number 0

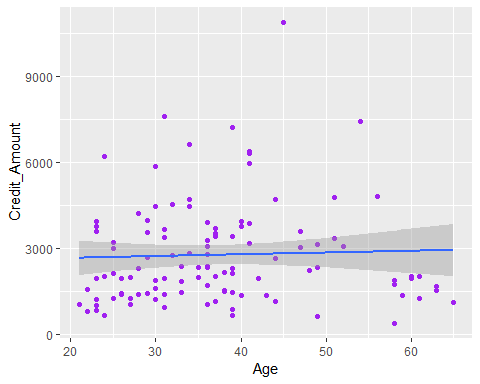


## `geomsmooth()` using method = 'loess' and formula 'y ~ x'  
  
ggplot(credit\_ds,aes(Age,Credit\_Amount))+geom\_point(colour="red")+geom\_smooth(span=1.25)

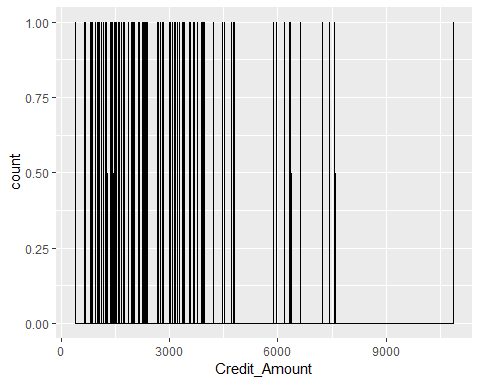
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'  
  
ggplot(credit\_ds,aes(Age,Credit\_Amount))+geom\_point(colour="purple")+geom\_smooth(method="lm")



#Facetting used to replicate the display for the two groups:  
# 1 Absence of Risk 2 Presence of Risk   
ggplot(credit\_ds,aes(Credit\_Amount,colour=Age))+geom\_freqpoly(binwidth=1)



ggplot(credit\_ds,aes(Credit\_Amount,colour=Age))+geom\_histogram(binwidth=.9)+facet\_wrap(~Age,ncol=1)



# Error: StatBin requires a continuous x variable: the x variable is discrete. Perhaps you want stat="count"?  
# Sum up the answer from the comments above:  
# 1 - Replace geom\_histogram(binwidth=0.5) with geom\_bar(). However this way will not allow binwidth customization.  
# 2 - Using stat\_count(width = 0.5) instead of geom\_bar() or geom\_histogram(binwidth = 0.5) would solve it.  
  
  
# ---------------------  
# ggplot(credit\_ds,aes(Credit\_Amount,colour="Yellow"))+geom\_bar()  
# ggplot(credit\_ds,aes(Credit\_Amount,colour=Age))+stat\_count(width = # 0.5)+facet\_wrap(~Creditability,ncol=2)