Challenge-7

Marzuki Nooranas

2023-10-02

**Practice:**

# Enter code here  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.3 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

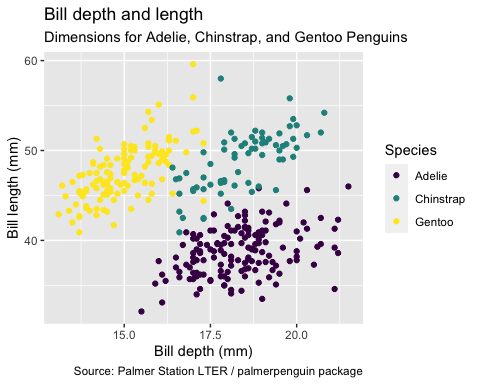
library(palmerpenguins)  
glimpse(penguins)

## Rows: 344  
## Columns: 8  
## $ species <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel…  
## $ island <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse…  
## $ bill\_length\_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, …  
## $ bill\_depth\_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, …  
## $ flipper\_length\_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186…  
## $ body\_mass\_g <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, …  
## $ sex <fct> male, female, female, NA, female, male, female, male…  
## $ year <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007…

**Plot recreation:**

ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 colour = species)) +   
 geom\_point() +  
 labs(title = "Bill depth and length",  
 subtitle = "Dimensions for Adelie, Chinstrap, and Gentoo Penguins",  
 x = "Bill depth (mm)",  
 y = "Bill length (mm)",  
 colour = "Species",  
 caption = "Source: Palmer Station LTER / palmerpenguin package") +  
scale\_colour\_viridis\_d()

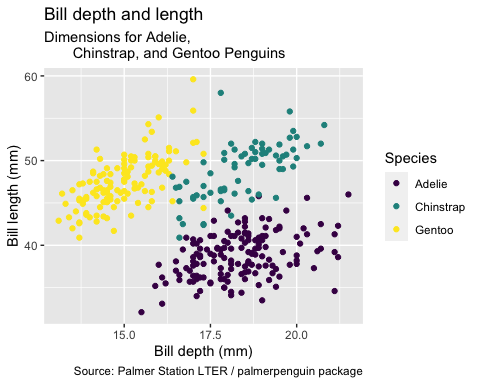
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Palmer Penguins: Argument names:**

ggplot(penguins,  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 colour= species)) +   
 geom\_point() +  
 labs(title = "Bill depth and length",  
 subtitle = "Dimensions for Adelie,  
 Chinstrap, and Gentoo Penguins",  
 x = "Bill depth (mm)",  
 y = "Bill length (mm)",  
 colour = "Species",  
 caption = "Source: Palmer Station LTER / palmerpenguin package") +  
scale\_colour\_viridis\_d()

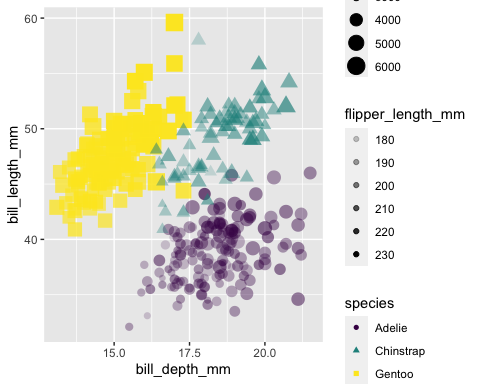
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Palmer Penguins:Aesthetics Options:**

ggplot(penguins, aes(x = bill\_depth\_mm, y = bill\_length\_mm, colour = species,  
shape = species, size = body\_mass\_g, alpha = flipper\_length\_mm)) +  
geom\_point() + scale\_colour\_viridis\_d()

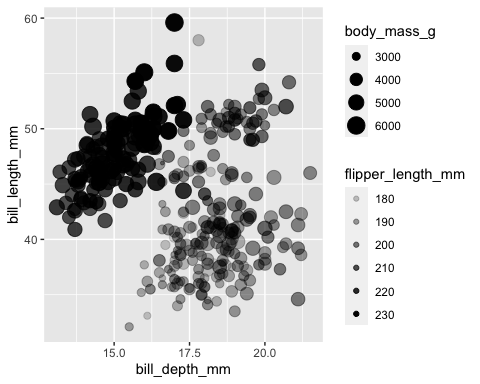
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Mapping:**

ggplot(penguins) +  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 size = body\_mass\_g,  
 alpha = flipper\_length\_mm) +  
geom\_point()

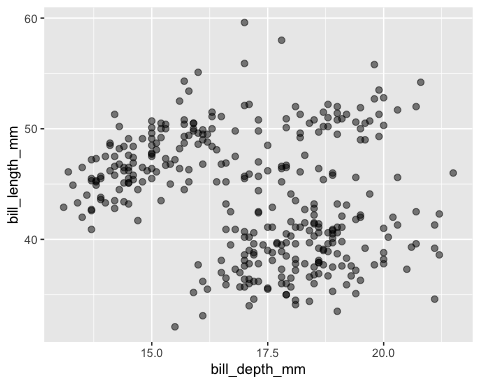
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Setting:**

ggplot(penguins) +  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm) +  
 geom\_point(size = 2, alpha = 0.5)

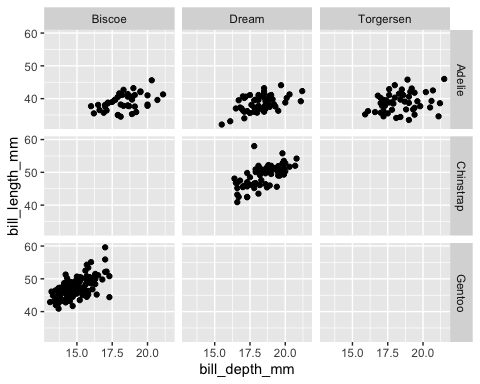
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Faceting:**

ggplot(penguins) +  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm) +  
 geom\_point() +  
 facet\_grid(species ~ island)

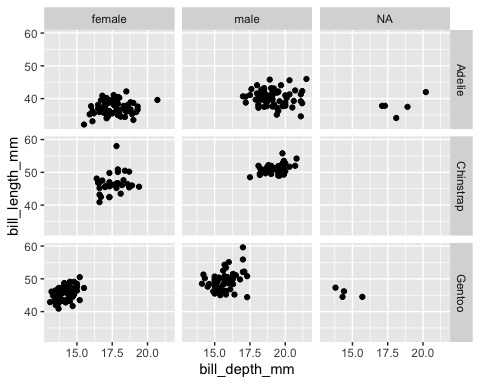
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Facet 2:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm)) +   
 geom\_point() +  
facet\_grid(species ~ sex)

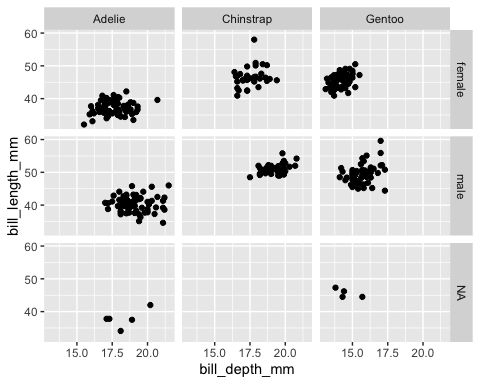
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Facet 3:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm)) +   
 geom\_point() +  
facet\_grid(sex ~ species)

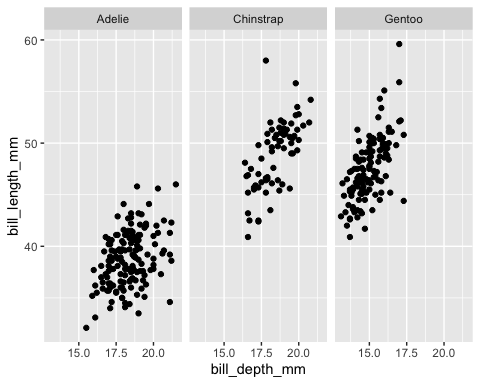
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Facet 4:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm)) +   
 geom\_point() +  
facet\_wrap(~ species)

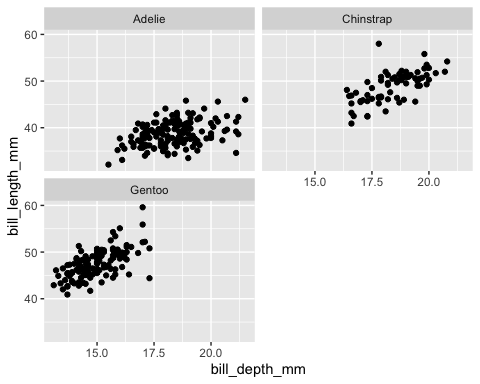
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Facet 5:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm)) +   
 geom\_point() +  
facet\_wrap(~ species, ncol = 2)

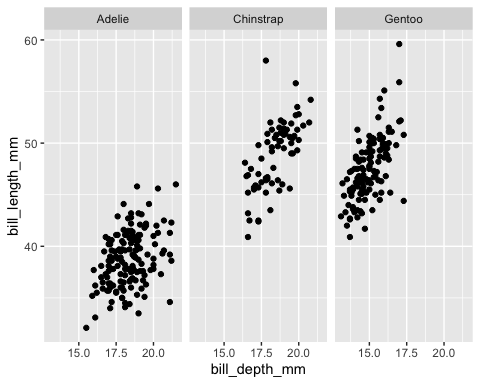
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Facet 6:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm)) +   
 geom\_point() +  
facet\_grid(. ~ species)

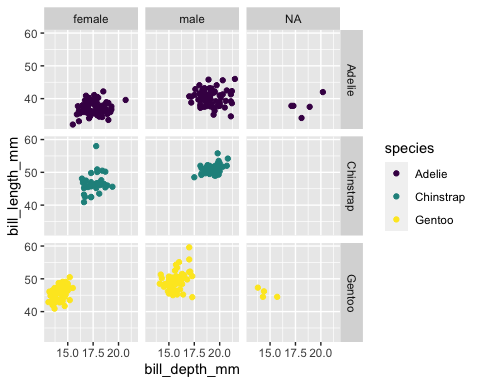
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Facet and Colour:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm,   
 color = species)) +  
geom\_point() +   
 facet\_grid(species ~ sex) +   
 scale\_color\_viridis\_d()

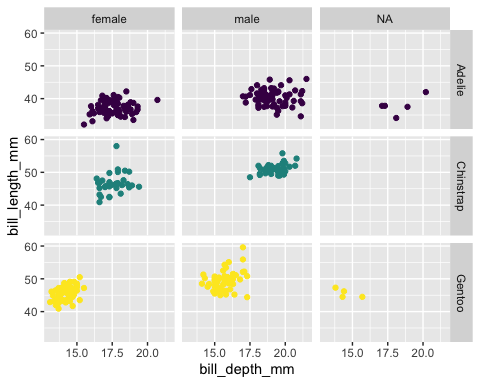
## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Face and color, no legend:**

ggplot(penguins,   
 aes(x = bill\_depth\_mm,   
 y = bill\_length\_mm,   
 color = species)) +  
geom\_point() +   
 facet\_grid(species ~ sex) +   
 scale\_color\_viridis\_d() +  
guides(color = "none")

## Warning: Removed 2 rows containing missing values (`geom\_point()`).



**Dataset 2: Lending Club**

library(openintro)

## Loading required package: airports

## Loading required package: cherryblossom

## Loading required package: usdata

glimpse(loans\_full\_schema)

## Rows: 10,000  
## Columns: 55  
## $ emp\_title <chr> "global config engineer ", "warehouse…  
## $ emp\_length <dbl> 3, 10, 3, 1, 10, NA, 10, 10, 10, 3, 1…  
## $ state <fct> NJ, HI, WI, PA, CA, KY, MI, AZ, NV, I…  
## $ homeownership <fct> MORTGAGE, RENT, RENT, RENT, RENT, OWN…  
## $ annual\_income <dbl> 90000, 40000, 40000, 30000, 35000, 34…  
## $ verified\_income <fct> Verified, Not Verified, Source Verifi…  
## $ debt\_to\_income <dbl> 18.01, 5.04, 21.15, 10.16, 57.96, 6.4…  
## $ annual\_income\_joint <dbl> NA, NA, NA, NA, 57000, NA, 155000, NA…  
## $ verification\_income\_joint <fct> , , , , Verified, , Not Verified, , ,…  
## $ debt\_to\_income\_joint <dbl> NA, NA, NA, NA, 37.66, NA, 13.12, NA,…  
## $ delinq\_2y <int> 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0…  
## $ months\_since\_last\_delinq <int> 38, NA, 28, NA, NA, 3, NA, 19, 18, NA…  
## $ earliest\_credit\_line <dbl> 2001, 1996, 2006, 2007, 2008, 1990, 2…  
## $ inquiries\_last\_12m <int> 6, 1, 4, 0, 7, 6, 1, 1, 3, 0, 4, 4, 8…  
## $ total\_credit\_lines <int> 28, 30, 31, 4, 22, 32, 12, 30, 35, 9,…  
## $ open\_credit\_lines <int> 10, 14, 10, 4, 16, 12, 10, 15, 21, 6,…  
## $ total\_credit\_limit <int> 70795, 28800, 24193, 25400, 69839, 42…  
## $ total\_credit\_utilized <int> 38767, 4321, 16000, 4997, 52722, 3898…  
## $ num\_collections\_last\_12m <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ num\_historical\_failed\_to\_pay <int> 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0…  
## $ months\_since\_90d\_late <int> 38, NA, 28, NA, NA, 60, NA, 71, 18, N…  
## $ current\_accounts\_delinq <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ total\_collection\_amount\_ever <int> 1250, 0, 432, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ current\_installment\_accounts <int> 2, 0, 1, 1, 1, 0, 2, 2, 6, 1, 2, 1, 2…  
## $ accounts\_opened\_24m <int> 5, 11, 13, 1, 6, 2, 1, 4, 10, 5, 6, 7…  
## $ months\_since\_last\_credit\_inquiry <int> 5, 8, 7, 15, 4, 5, 9, 7, 4, 17, 3, 4,…  
## $ num\_satisfactory\_accounts <int> 10, 14, 10, 4, 16, 12, 10, 15, 21, 6,…  
## $ num\_accounts\_120d\_past\_due <int> 0, 0, 0, 0, 0, 0, 0, NA, 0, 0, 0, 0, …  
## $ num\_accounts\_30d\_past\_due <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ num\_active\_debit\_accounts <int> 2, 3, 3, 2, 10, 1, 3, 5, 11, 3, 2, 2,…  
## $ total\_debit\_limit <int> 11100, 16500, 4300, 19400, 32700, 272…  
## $ num\_total\_cc\_accounts <int> 14, 24, 14, 3, 20, 27, 8, 16, 19, 7, …  
## $ num\_open\_cc\_accounts <int> 8, 14, 8, 3, 15, 12, 7, 12, 14, 5, 8,…  
## $ num\_cc\_carrying\_balance <int> 6, 4, 6, 2, 13, 5, 6, 10, 14, 3, 5, 3…  
## $ num\_mort\_accounts <int> 1, 0, 0, 0, 0, 3, 2, 7, 2, 0, 2, 3, 3…  
## $ account\_never\_delinq\_percent <dbl> 92.9, 100.0, 93.5, 100.0, 100.0, 78.1…  
## $ tax\_liens <int> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ public\_record\_bankrupt <int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0…  
## $ loan\_purpose <fct> moving, debt\_consolidation, other, de…  
## $ application\_type <fct> individual, individual, individual, i…  
## $ loan\_amount <int> 28000, 5000, 2000, 21600, 23000, 5000…  
## $ term <dbl> 60, 36, 36, 36, 36, 36, 60, 60, 36, 3…  
## $ interest\_rate <dbl> 14.07, 12.61, 17.09, 6.72, 14.07, 6.7…  
## $ installment <dbl> 652.53, 167.54, 71.40, 664.19, 786.87…  
## $ grade <fct> C, C, D, A, C, A, C, B, C, A, C, B, C…  
## $ sub\_grade <fct> C3, C1, D1, A3, C3, A3, C2, B5, C2, A…  
## $ issue\_month <fct> Mar-2018, Feb-2018, Feb-2018, Jan-201…  
## $ loan\_status <fct> Current, Current, Current, Current, C…  
## $ initial\_listing\_status <fct> whole, whole, fractional, whole, whol…  
## $ disbursement\_method <fct> Cash, Cash, Cash, Cash, Cash, Cash, C…  
## $ balance <dbl> 27015.86, 4651.37, 1824.63, 18853.26,…  
## $ paid\_total <dbl> 1999.330, 499.120, 281.800, 3312.890,…  
## $ paid\_principal <dbl> 984.14, 348.63, 175.37, 2746.74, 1569…  
## $ paid\_interest <dbl> 1015.19, 150.49, 106.43, 566.15, 754.…  
## $ paid\_late\_fees <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…

**Selected Variable:**

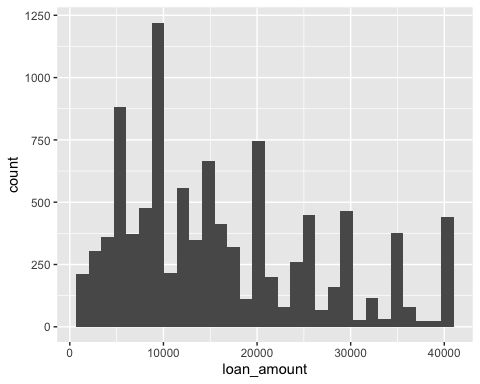
loans <- loans\_full\_schema %>%  
 select(loan\_amount, interest\_rate, term, grade,  
 state, annual\_income, homeownership, debt\_to\_income)  
glimpse(loans)

## Rows: 10,000  
## Columns: 8  
## $ loan\_amount <int> 28000, 5000, 2000, 21600, 23000, 5000, 24000, 20000, 20…  
## $ interest\_rate <dbl> 14.07, 12.61, 17.09, 6.72, 14.07, 6.72, 13.59, 11.99, 1…  
## $ term <dbl> 60, 36, 36, 36, 36, 36, 60, 60, 36, 36, 60, 60, 36, 60,…  
## $ grade <fct> C, C, D, A, C, A, C, B, C, A, C, B, C, B, D, D, D, F, E…  
## $ state <fct> NJ, HI, WI, PA, CA, KY, MI, AZ, NV, IL, IL, FL, SC, CO,…  
## $ annual\_income <dbl> 90000, 40000, 40000, 30000, 35000, 34000, 35000, 110000…  
## $ homeownership <fct> MORTGAGE, RENT, RENT, RENT, RENT, OWN, MORTGAGE, MORTGA…  
## $ debt\_to\_income <dbl> 18.01, 5.04, 21.15, 10.16, 57.96, 6.46, 23.66, 16.19, 3…

**Histogram**

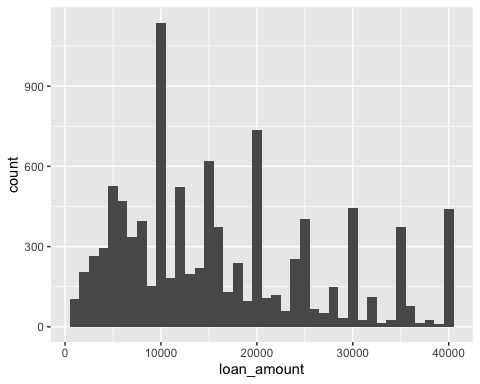
ggplot(loans) +   
 aes(x = loan\_amount) +  
geom\_histogram()

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



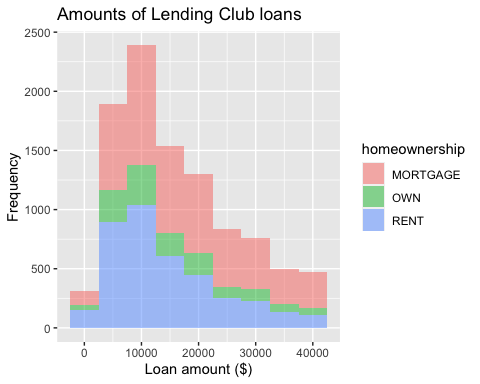
**Histograms and binwidth=1000:**

# binwidth = 1000  
ggplot(loans,   
 aes(x = loan\_amount)) +  
geom\_histogram(binwidth = 1000)



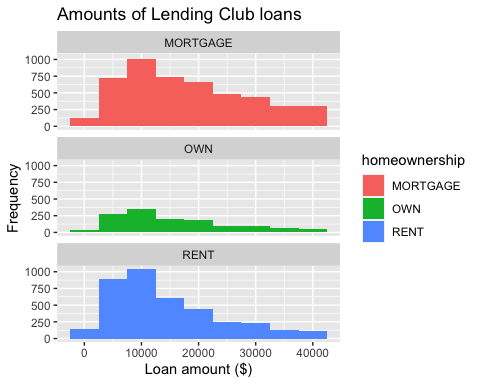
**Customising Histogram**

ggplot(loans,   
 aes(x = loan\_amount,   
 fill = homeownership)) +  
geom\_histogram(binwidth = 5000, alpha = 0.5) +  
labs(x = "Loan amount ($)",  
 y = "Frequency",  
 title = "Amounts of Lending Club loans")



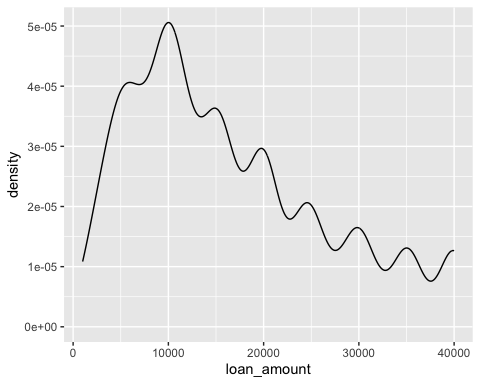
\*\*Facet with a categorical Variable:

ggplot(loans,   
 aes(x = loan\_amount, fill = homeownership)) +   
 geom\_histogram(binwidth = 5000) +  
labs(x = "Loan amount ($)",y = "Frequency",title = "Amounts of Lending Club loans") +  
facet\_wrap(~ homeownership, nrow = 3)



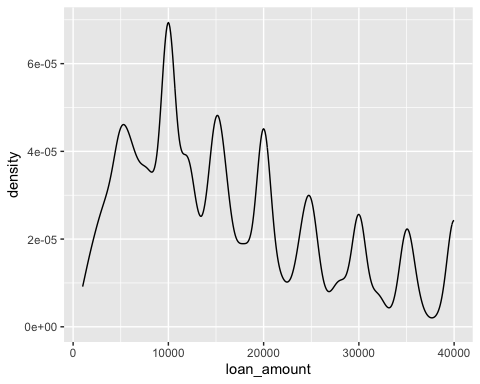
**Density Plot:**

ggplot(loans, aes(x = loan\_amount)) +  
geom\_density()



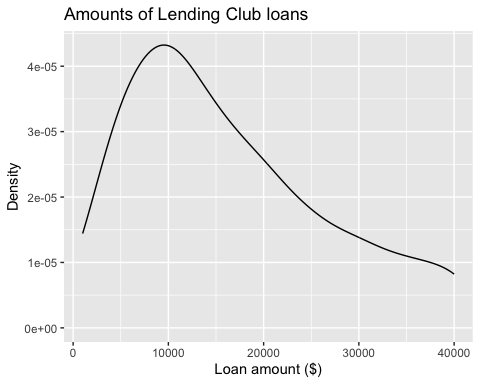
**Density Plot and adjusting bandwidth:**

ggplot(loans, aes(x = loan\_amount)) +  
geom\_density(adjust = 0.5)

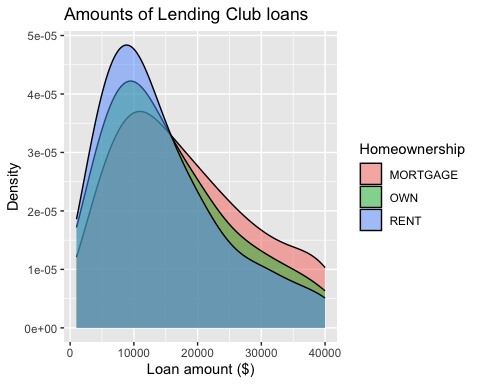


**Customising density plots:**

ggplot(loans, aes(x = loan\_amount)) +  
geom\_density(adjust = 2) +  
labs( x = "Loan amount ($)", y = "Density", title = "Amounts of Lending Club loans" )

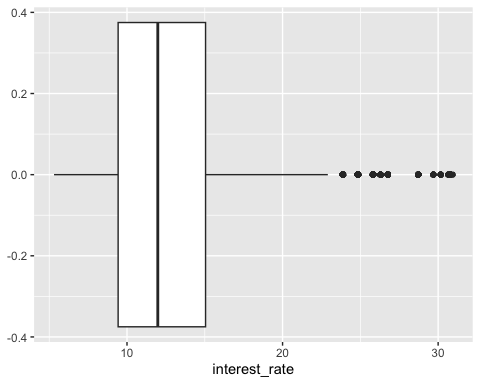


ggplot(loans, aes(x = loan\_amount, fill = homeownership)) +  
geom\_density(adjust = 2, alpha = 0.5) +  
labs(x = "Loan amount ($)",y = "Density",title = "Amounts of Lending Club loans", fill = "Homeownership")



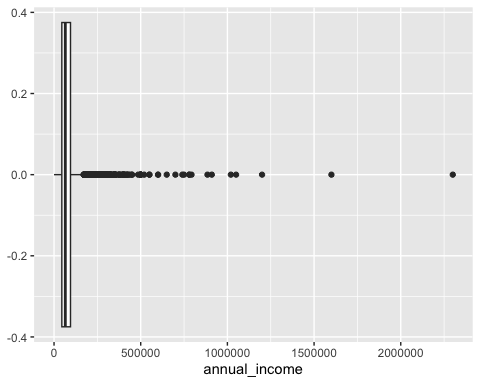
**Box plots:**

ggplot(loans, aes(x = interest\_rate)) +  
geom\_boxplot()



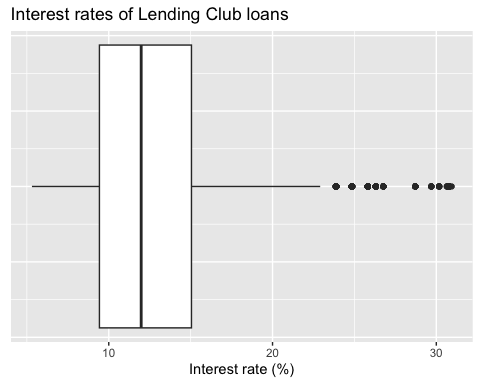
**Box plots and outliers:**

ggplot(loans, aes(x = annual\_income)) +  
geom\_boxplot()



**Customising Box Plots:**

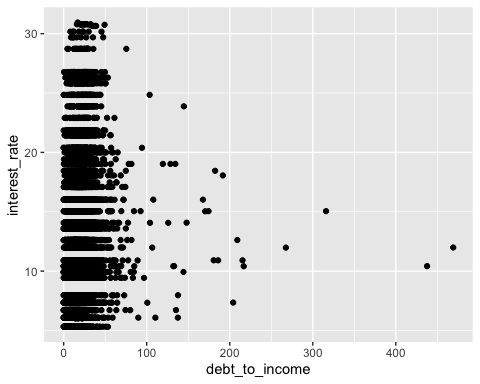
ggplot(loans,   
 aes(x = interest\_rate)) +  
 geom\_boxplot() +labs(x = "Interest rate (%)",  
 y = NULL,  
title = "Interest rates of Lending Club loans") +  
theme( axis.ticks.y = element\_blank(), axis.text.y = element\_blank() )



**Scatterplot:**

ggplot(loans,   
 aes(x = debt\_to\_income, y = interest\_rate)) +  
geom\_point()

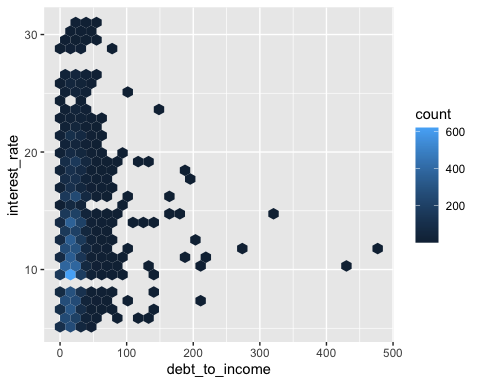
## Warning: Removed 24 rows containing missing values (`geom\_point()`).



**Hex plot:**

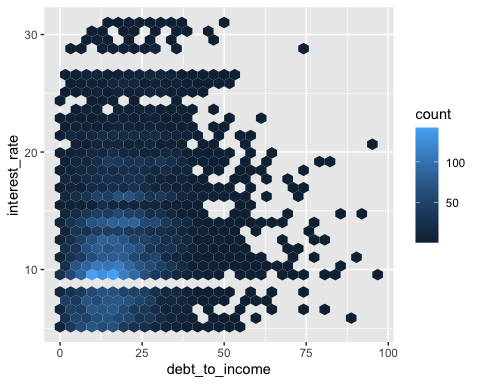
library(hexbin)  
ggplot(loans, aes(x = debt\_to\_income, y = interest\_rate)) +  
geom\_hex()

## Warning: Removed 24 rows containing non-finite values (`stat\_binhex()`).



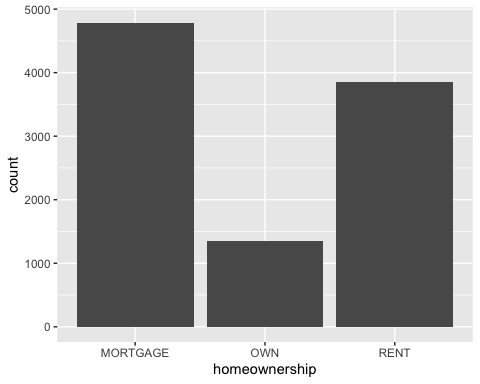
**Hex plot 2:**

ggplot(loans %>%   
 filter(debt\_to\_income < 100),  
aes(x = debt\_to\_income,   
 y = interest\_rate)) +  
geom\_hex()



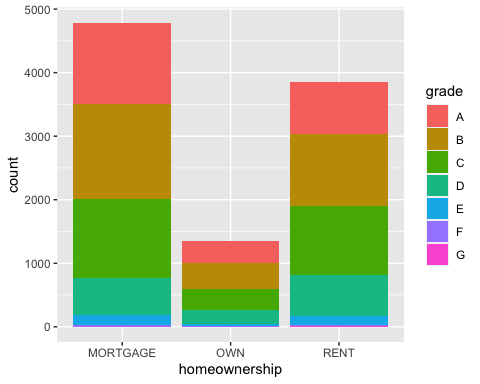
**Bar plot**

ggplot(loans, aes(x = homeownership)) +  
geom\_bar()

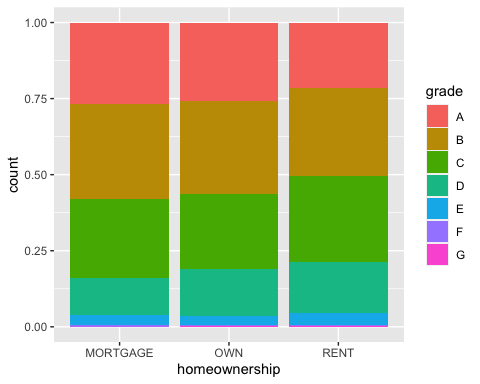


**Segmented bar plot:**

ggplot(loans,   
 aes(x = homeownership,  
 fill = grade)) +  
geom\_bar()

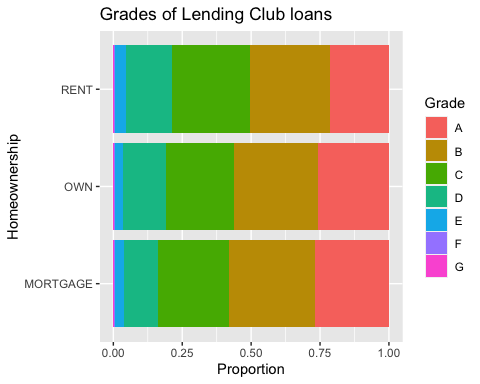


ggplot(loans, aes(x = homeownership,   
 fill = grade)) +  
geom\_bar(position = "fill")



**Customising Bar plots**

ggplot(loans, aes(y = homeownership, fill = grade)) + geom\_bar(position = "fill") +  
labs( x = "Proportion",   
 y = "Homeownership", fill = "Grade",   
 title = "Grades of Lending Club loans")



**Violin plots**

ggplot(loans,   
 aes(x = homeownership, y = loan\_amount)) +  
geom\_violin()



**Ridge plots:**

library(ggridges)  
ggplot(loans, aes(x = loan\_amount,   
 y = grade,   
 fill = grade,   
 color = grade)) +  
geom\_density\_ridges(alpha = 0.5)

## Picking joint bandwidth of 2360

