Week4Lecture

Zain

9/19/2021

## Tidyverse

tidyverse package contains the ggplot library that we will use for graph. So, we will import tidyverse first

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

###Importing data Let us import the dataset

mn\_homes <- read\_csv("data/mn\_homes.csv")

## Rows: 495 Columns: 13

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): neighborhood, community  
## dbl (10): saleyear, salemonth, salesprice, area, beds, baths, stories, yearb...  
## lgl (1): fireplace

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

glimpse(mn\_homes)

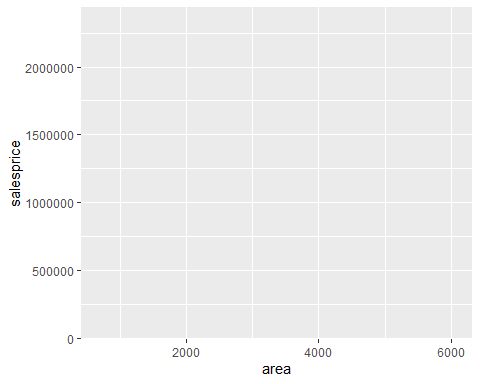
## Rows: 495  
## Columns: 13  
## $ saleyear <dbl> 2012, 2014, 2005, 2010, 2010, 2013, 2011, 2007, 2013, 20~  
## $ salemonth <dbl> 6, 7, 7, 6, 2, 9, 1, 9, 10, 6, 7, 8, 5, 2, 7, 6, 10, 6, ~  
## $ salesprice <dbl> 690467.0, 235571.7, 272507.7, 277767.5, 148324.1, 242871~  
## $ area <dbl> 3937, 1440, 1835, 2016, 2004, 2822, 2882, 1979, 3140, 35~  
## $ beds <dbl> 5, 2, 2, 3, 3, 3, 4, 3, 4, 3, 3, 3, 2, 3, 3, 6, 2, 3, 2,~  
## $ baths <dbl> 4, 1, 1, 2, 1, 3, 3, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 1,~  
## $ stories <dbl> 2.5, 1.7, 1.7, 2.5, 1.0, 2.0, 1.7, 1.5, 1.5, 2.5, 1.0, 2~  
## $ yearbuilt <dbl> 1907, 1919, 1913, 1910, 1956, 1934, 1951, 1929, 1940, 19~  
## $ neighborhood <chr> "Lowry Hill", "Cooper", "Hiawatha", "King Field", "Shing~  
## $ community <chr> "Calhoun-Isles", "Longfellow", "Longfellow", "Southwest"~  
## $ lotsize <dbl> 6192, 5160, 5040, 4875, 5060, 6307, 6500, 5600, 6350, 75~  
## $ numfireplaces <dbl> 0, 0, 0, 0, 0, 2, 2, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0,~  
## $ fireplace <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, TRUE, FALSE, TR~

summary(mn\_homes)

## saleyear salemonth salesprice area   
## Min. :2005 Min. : 1.000 Min. : 103526 Min. : 682   
## 1st Qu.:2007 1st Qu.: 4.000 1st Qu.: 195327 1st Qu.:1783   
## Median :2010 Median : 7.000 Median : 256564 Median :2121   
## Mean :2010 Mean : 6.709 Mean : 305759 Mean :2212   
## 3rd Qu.:2013 3rd Qu.: 9.000 3rd Qu.: 338242 3rd Qu.:2493   
## Max. :2015 Max. :12.000 Max. :2328760 Max. :6053   
## beds baths stories yearbuilt   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1886   
## 1st Qu.:3.000 1st Qu.:1.000 1st Qu.:1.200 1st Qu.:1914   
## Median :3.000 Median :2.000 Median :1.500 Median :1925   
## Mean :3.087 Mean :1.865 Mean :1.482 Mean :1931   
## 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:1.700 3rd Qu.:1947   
## Max. :7.000 Max. :6.000 Max. :2.500 Max. :2015   
## neighborhood community lotsize numfireplaces   
## Length:495 Length:495 Min. : 2081 Min. :0.0000   
## Class :character Class :character 1st Qu.: 5080 1st Qu.:0.0000   
## Mode :character Mode :character Median : 5359 Median :0.0000   
## Mean : 5872 Mean :0.4949   
## 3rd Qu.: 6322 3rd Qu.:1.0000   
## Max. :21312 Max. :3.0000   
## fireplace   
## Mode :logical   
## FALSE:304   
## TRUE :191   
##   
##   
##

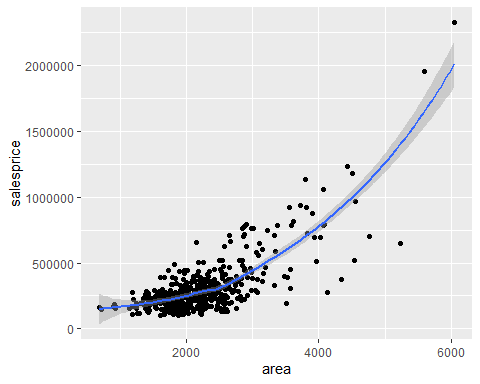
###first ggplot :) ggplot just give you the basic background of the plot

ggplot(data = mn\_homes,mapping = aes(x=area,y=salesprice))

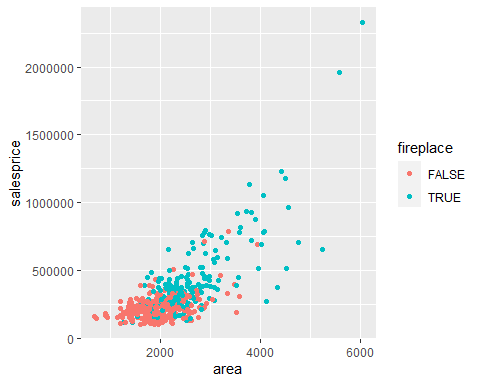
 Now, let us add the actual values

ggplot(data = mn\_homes,mapping = aes(x=area,y=salesprice))+  
 geom\_point()+  
 geom\_smooth()

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

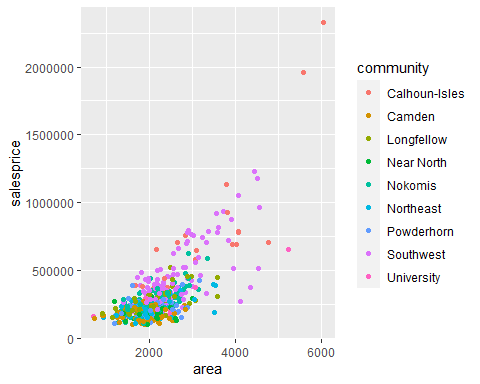
 Now, let us draw three variables in our diagram

ggplot(data = mn\_homes,mapping = aes(x=area,y=salesprice,color=fireplace))+  
 geom\_point()

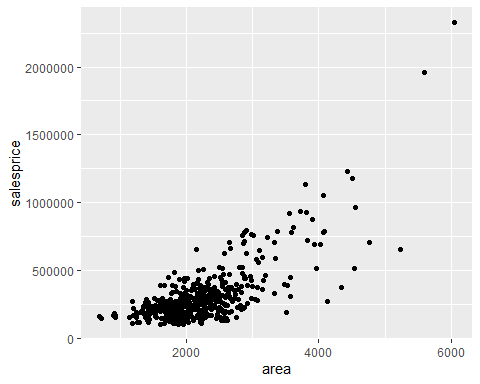


different variable for color

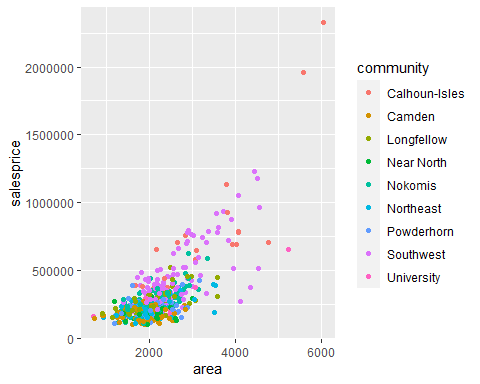
ggplot(data = mn\_homes,mapping = aes(x=area,y=salesprice,color=community))+  
 geom\_point()

 Quick Revision for the last session

ggplot(data=mn\_homes, mapping = aes(x=area,y=salesprice)) +  
 geom\_point()

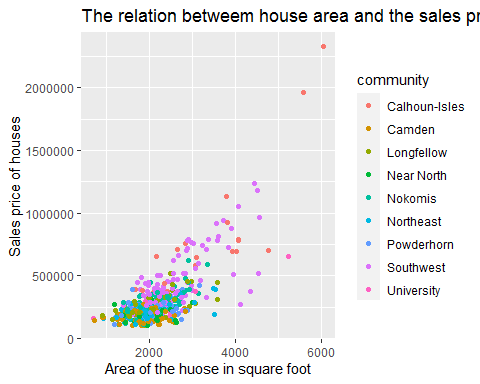


ggplot(data=mn\_homes, mapping = aes(x=area,y=salesprice, color=community)) +  
 geom\_point()

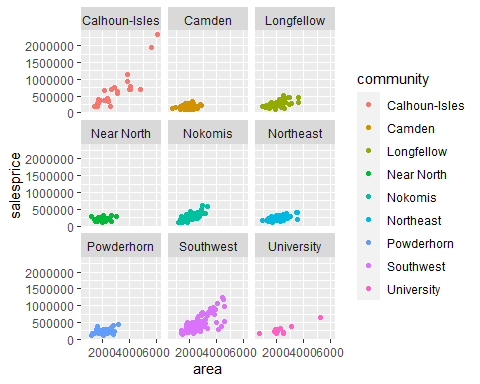


### Adding lables to the graph

ggplot(data=mn\_homes, mapping = aes(x=area,y=salesprice, color=community)) +  
 geom\_point() +  
 labs(title="The relation betweem house area and the sales price (2005-2015)", x="Area of the huose in square foot", y="Sales price of houses")

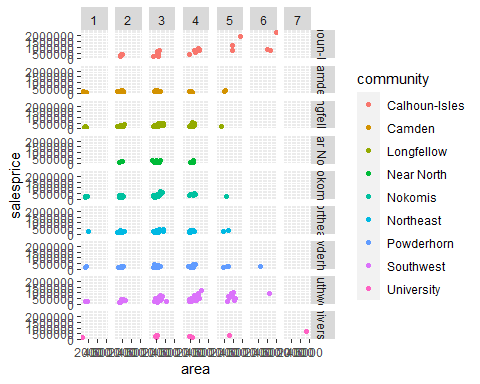
 ##Faceting Sometimes we get too much points in our graph. We can use faceting to seperate them in several graphs. for example:

ggplot(data=mn\_homes, mapping = aes(x=area,y=salesprice, color=community)) +  
 geom\_point() +  
 facet\_wrap(~community)



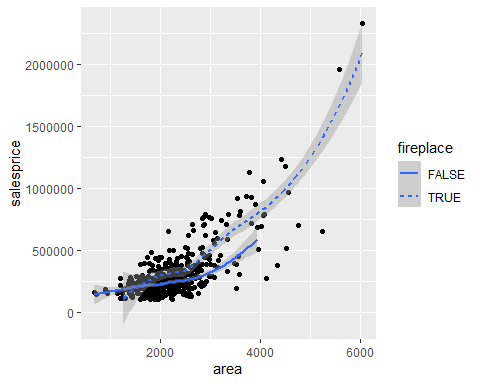
The other type of faceting is facet\_grid(). It help you to categorize your charts based on two variables. For example:

ggplot(data=mn\_homes, mapping = aes(x=area,y=salesprice, color=community)) +  
 geom\_point() +  
 facet\_grid(community~beds)

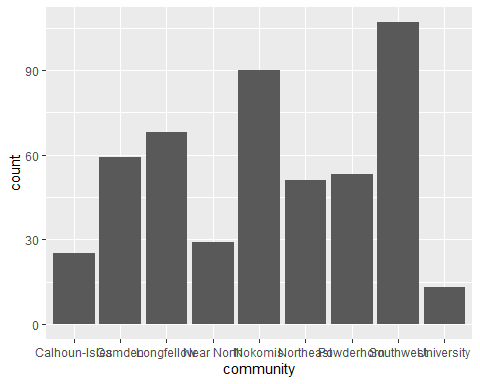
 let us get back to geom\_smooth and check the linetype attribute

ggplot(data=mn\_homes, mapping = aes(x=area, y=salesprice, linetype=fireplace))+  
 geom\_point()+  
 geom\_smooth()

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

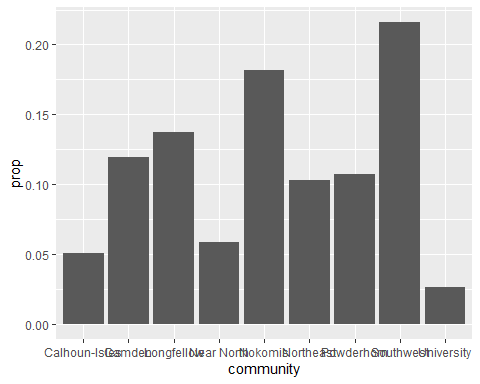
 ##geom\_bar() We use geom\_bar with discrete variables. It ussually use on variable only

ggplot(data=mn\_homes) +  
 geom\_bar(mapping=aes(x=community))



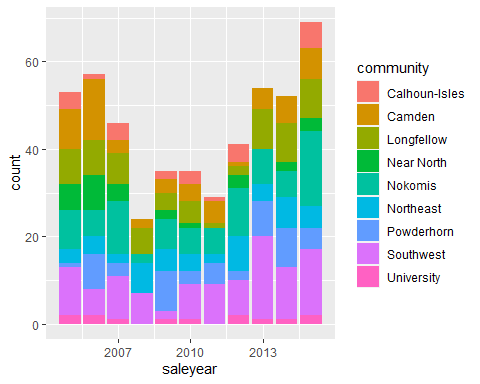
count is the default y axis for the bar chart. However, we can change this manually to percentage. for example:

ggplot(data=mn\_homes) +  
 geom\_bar(mapping=aes(x=community, y=..prop.., group=1))

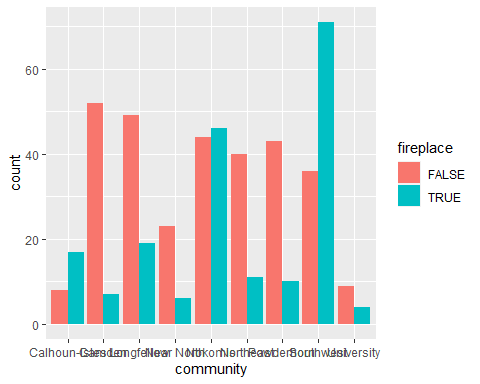


We can add more variables to bar charts using color or fill attributes

ggplot(data=mn\_homes) +  
 geom\_bar(mapping=aes(x=saleyear,fill=community))

 ### poistion attribute for bar chart

ggplot(data=mn\_homes) +  
 geom\_bar(mapping=aes(x=community,fill=fireplace),position="dodge")



ggplot(data=mn\_homes) +  
 geom\_bar(mapping=aes(x=community,fill=fireplace),position="identity", alpha=1/5)

