library("tidyverse")

## Warning: package 'tidyverse' was built under R version 4.1.3

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

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.5 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.0.2 v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.1.3

## Warning: package 'dplyr' was built under R version 4.1.2

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

library("ggplot2")  
library("psych")

## Warning: package 'psych' was built under R version 4.1.2

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library("dplyr")  
file <- read\_csv("datasets/ethnic data.csv")

## New names:  
## \* Emergency -> Emergency...7  
## \* Emergency -> Emergency...42

## Rows: 18 Columns: 44

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (1): Ethnic category  
## dbl (5): Mean time waited, Median time waited, Mean length of stay, Median l...

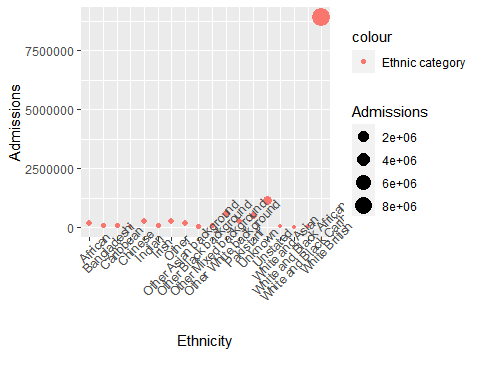
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

file

## # A tibble: 18 x 44  
## `Ethnic category` `Finished consul~ Admissions Male Female `Gender Unknown`  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 White British 11451204 8896629 5.21e6 6.23e6 11569  
## 2 Irish 114463 85844 5.27e4 6.13e4 393  
## 3 Other White back~ 732353 609632 3.00e5 4.32e5 1012  
## 4 White and Black ~ 49128 42908 1.84e4 3.07e4 105  
## 5 White and Black ~ 24348 21240 1.00e4 1.43e4 54  
## 6 White and Asian 38179 33744 1.71e4 2.10e4 151  
## 7 Other Mixed back~ 82254 71402 3.59e4 4.62e4 147  
## 8 Indian 296389 241271 1.30e5 1.65e5 1173  
## 9 Pakistani 290749 246010 1.20e5 1.71e5 565  
## 10 Bangladeshi 88271 73948 3.74e4 5.08e4 111  
## 11 Other Asian back~ 205084 167960 9.36e4 1.11e5 636  
## 12 Caribbean 144954 109880 6.03e4 8.44e4 256  
## 13 African 212261 175873 8.37e4 1.28e5 438  
## 14 Other Black back~ 86653 69018 3.81e4 4.84e4 91  
## 15 Chinese 36473 29998 1.48e4 2.16e4 79  
## 16 Other 329291 267695 1.50e5 1.79e5 364  
## 17 Unknown 610571 515939 2.99e5 3.08e5 3864  
## 18 Unstated 1376064 1154129 6.53e5 7.14e5 8311  
## # ... with 38 more variables: Emergency...7 <dbl>, Waiting list <dbl>,  
## # Planned <dbl>, Other Admission Method <dbl>, Mean time waited <dbl>,  
## # Median time waited <dbl>, Mean length of stay <dbl>,  
## # Median length of stay <dbl>, 8 <dbl>, Age 0 <dbl>, Age 1-4 <dbl>,  
## # Age 5-9 <dbl>, Age 10-14 <dbl>, Age 15 <dbl>, Age 16 <dbl>, Age 17 <dbl>,  
## # Age 18 <dbl>, Age 19 <dbl>, Age 20-24 <dbl>, Age 25-29 <dbl>,  
## # Age 30-34 <dbl>, Age 35-39 <dbl>, Age 40-44 <dbl>, Age 45-49 <dbl>, ...

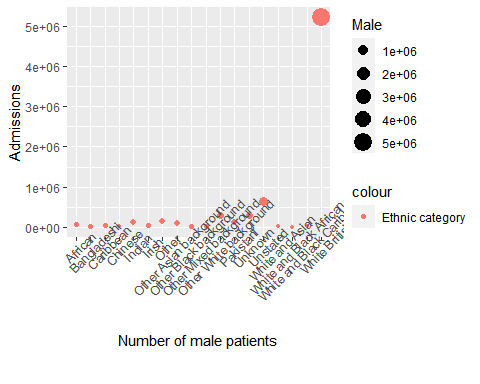
file2 <- file %>% select("Ethnic category", Admissions, Male, Female)  
ggplot(file2, aes(file$`Ethnic category`, file$Admissions)) +geom\_point((aes(col='Ethnic category', size=Admissions))) +geom\_smooth()+theme(axis.text.x = element\_text(angle = 45)) + theme(axis.title.x = element\_text(vjust=10.5)) + ylab("Admissions") +xlab("Ethnicity")

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



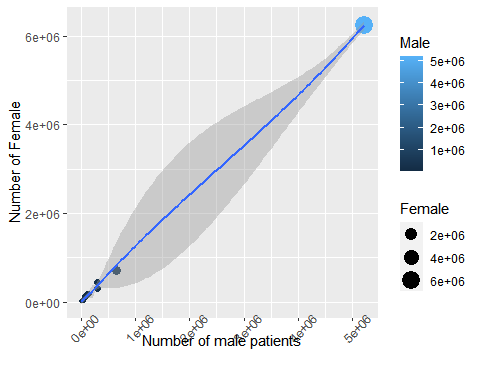
# The White British community is over represented due to the fact that they are the majority of the population, so it is something to take note of when analyzing  
  
#comparing Ethnic category and the male admissions number  
ggplot(file2, aes(file$`Ethnic category`, file$Male)) +geom\_point((aes(col='Ethnic category', size=Male))) +geom\_smooth()+theme(axis.text.x = element\_text(angle = 45)) + theme(axis.title.x = element\_text(vjust=10.5)) + ylab("Admissions") +xlab("Number of male patients")

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



#comparing Female and the Male admissions number  
ggplot(file2, aes(file$Male, file$Female)) +geom\_point((aes(col=Male, size=Female))) +geom\_smooth()+theme(axis.text.x = element\_text(angle = 45)) + theme(axis.title.x = element\_text(vjust=10.5)) + ylab("Number of Female") +xlab("Number of male patients")

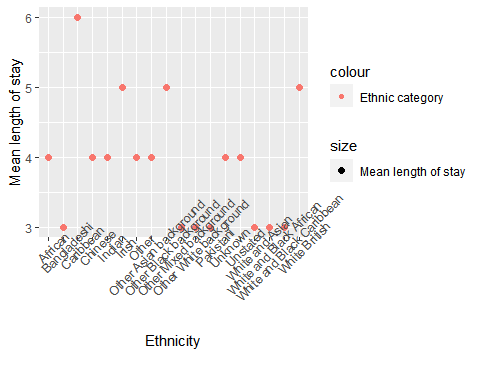
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



#comparing Ethnic category and the mean length of stay  
file3 <- file %>% select("Ethnic category", Admissions, "Mean length of stay")  
ggplot(file3, aes(file$`Ethnic category`, file$"Mean length of stay")) +geom\_point((aes(col='Ethnic category', size='Mean length of stay'))) +geom\_smooth()+theme(axis.text.x = element\_text(angle = 45)) + theme(axis.title.x = element\_text(vjust=10.5)) + ylab("Mean length of stay") +xlab("Ethnicity")

## Warning: Using size for a discrete variable is not advised.

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



#correlation testing  
library("ggpubr")

## Warning: package 'ggpubr' was built under R version 4.1.3

cor.test(file$Male, file$Female, method=c("pearson", "kendall", "spearman"))

##   
## Pearson's product-moment correlation  
##   
## data: file$Male and file$Female  
## t = 201.07, df = 16, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9994558 0.9999281  
## sample estimates:  
## cor   
## 0.9998022

cor.test(file3$`Admissions`, file$"Mean length of stay", method=c("pearson", "kendall", "spearman"))

##   
## Pearson's product-moment correlation  
##   
## data: file3$Admissions and file$"Mean length of stay"  
## t = 1.2767, df = 16, p-value = 0.2199  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1897371 0.6751007  
## sample estimates:  
## cor   
## 0.3040681

library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

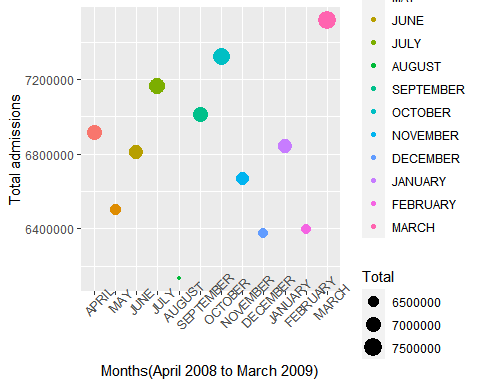
## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2008-09"  
timeseries\_subset <- timeseries[c(1:12), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
  
  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2008 to March 2009)") + geom\_path()

## geom\_path: Each group consists of only one observation. Do you need to adjust  
## the group aesthetic?



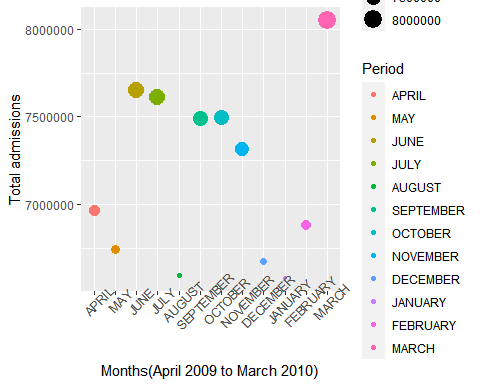
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2009-10"  
timeseries\_subset <- timeseries[c(13:24), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2009 to March 2010)")



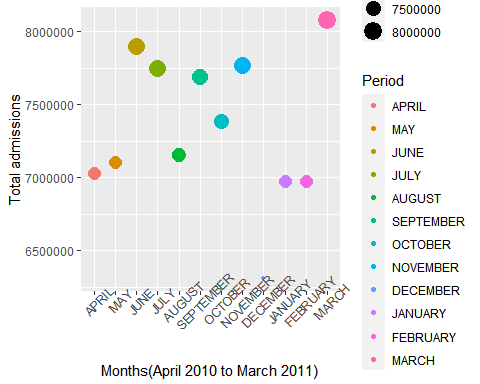
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2010-11"  
timeseries\_subset <- timeseries[c(25:36), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2010 to March 2011)")



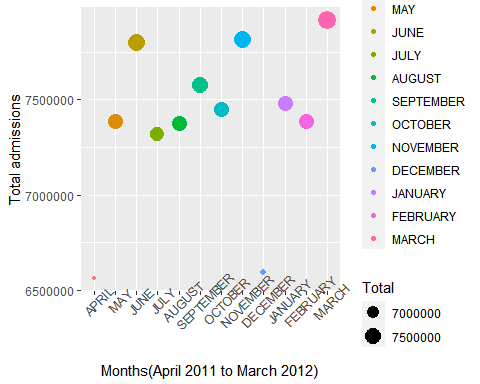
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2011-12"  
timeseries\_subset <- timeseries[c(37:48), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2011 to March 2012)")



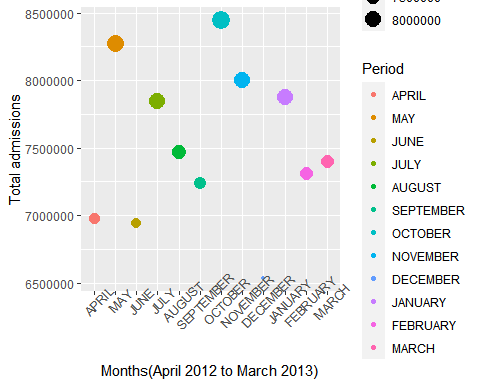
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2012-13"  
timeseries\_subset <- timeseries[c(49:60), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2012 to March 2013)")



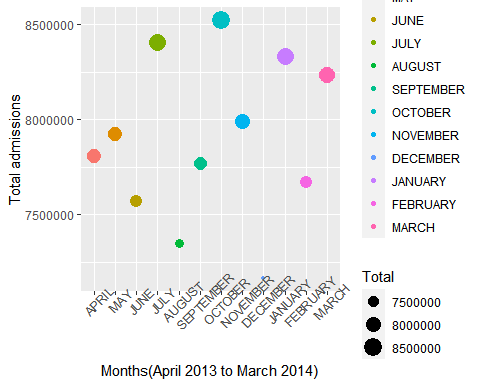
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2013-14"  
timeseries\_subset <- timeseries[c(61:72), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2013 to March 2014)")



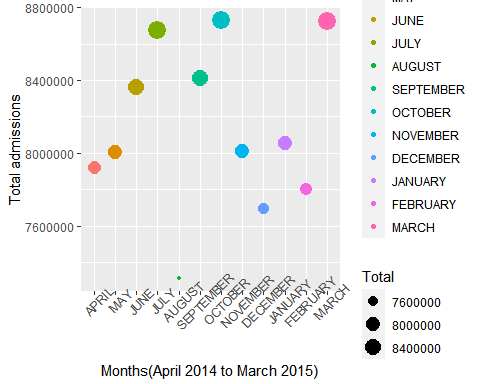
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2014-15"  
timeseries\_subset <- timeseries[c(73:84), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2014 to March 2015)")



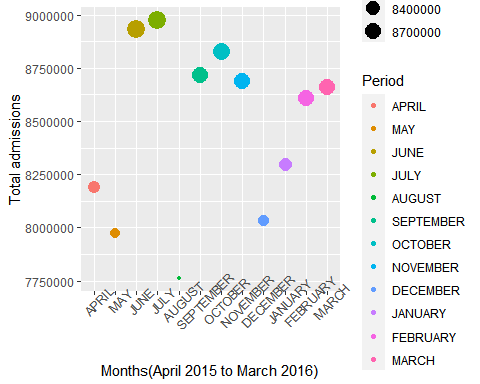
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2015-16"  
timeseries\_subset <- timeseries[c(85:96), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2015 to March 2016)")



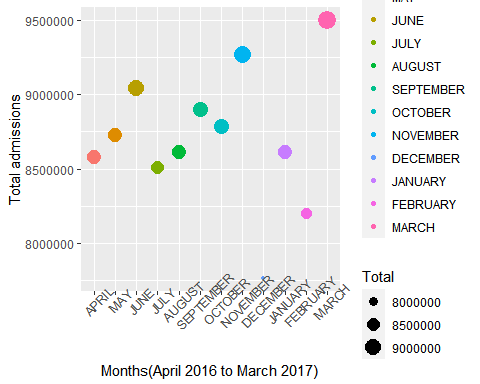
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2016-17"  
timeseries\_subset <- timeseries[c(97:108), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2016 to March 2017)")



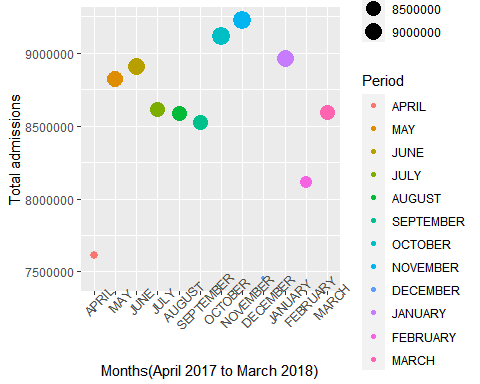
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

#"subset for 2017-18"  
timeseries\_subset <- timeseries[c(109:120), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
ggplot(timeseries\_subset, aes(Period, Total))+ geom\_point((aes(col=Period, size=Total))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Total admissions") +xlab("Months(April 2017 to March 2018)")



library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/10 year monthly summarized data.csv")

## Rows: 121 Columns: 17

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period

##   
## 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.

subtime <- timeseries[c(1:96), c(1,2,17)]  
  
timeseries\_subset$Period<-factor(timeseries\_subset$Period, levels = timeseries\_subset$Period)  
library("ggpubr")  
cor.test(file$Male, file$Female, method=c("pearson", "kendall", "spearman"))

##   
## Pearson's product-moment correlation  
##   
## data: file$Male and file$Female  
## t = 201.07, df = 16, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9994558 0.9999281  
## sample estimates:  
## cor   
## 0.9998022

cor.test(file3$`Admissions`, file$"Mean length of stay", method=c("pearson", "kendall", "spearman"))

##   
## Pearson's product-moment correlation  
##   
## data: file3$Admissions and file$"Mean length of stay"  
## t = 1.2767, df = 16, p-value = 0.2199  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1897371 0.6751007  
## sample estimates:  
## cor   
## 0.3040681

library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
timeseries <- read\_csv("datasets/new.csv")

## Rows: 120 Columns: 3

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Year, Period  
## dbl (1): Total

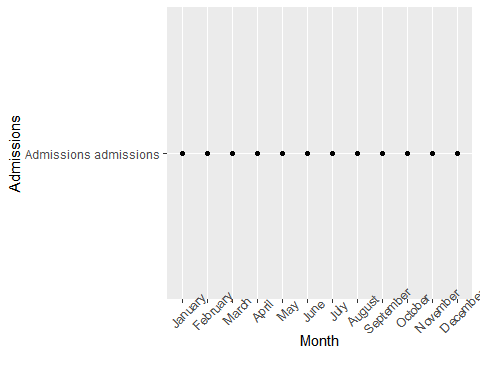
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

averages <- data.frame(Month = c("January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November","December"),  
 `Admissions admissions` = c(7800428.1, 7535116.6, 8268049.3, 7456169.3, 7745642.2, 7992003.2, 8088004.2, 7434678.1, 7932904.1, 8208671.9, 8074650.7, 7060764.5))  
  
averages$Month<-factor(averages$Month, levels = averages$Month)  
  
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
  
sapply(averages, class)

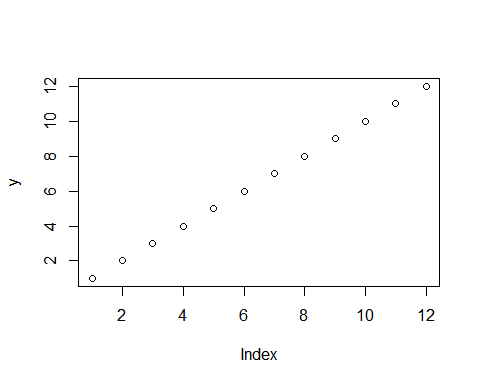
## Month Admissions.admissions   
## "factor" "numeric"

ggplot(averages, aes(Month,"Admissions admissions")) +geom\_point() +geom\_smooth()+theme(axis.text.x = element\_text(angle = 45)) + theme(axis.title.x = element\_text(vjust=10.5)) + ylab("Admissions") +xlab("Month") + geom\_smooth()

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



plot(averages$Month, averages$`Admissions admissions`)



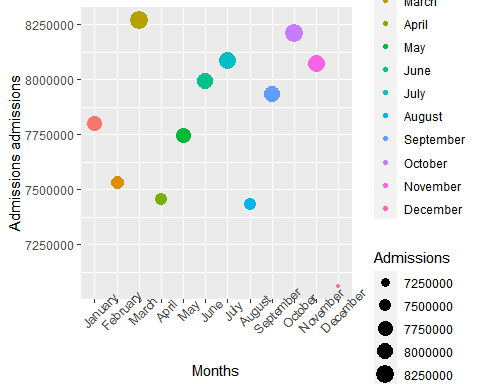
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
  
averages<- read\_csv("datasets/average admissions.csv")

## Rows: 12 Columns: 3

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (1): Month  
## dbl (2): Admissions, average maternity

##   
## 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.

averages$Month<-factor(averages$Month, levels = averages$Month)  
  
ggplot(averages, aes(Month, Admissions)) + geom\_point((aes(col=Month, size=Admissions))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Admissions admissions") +xlab("Months")



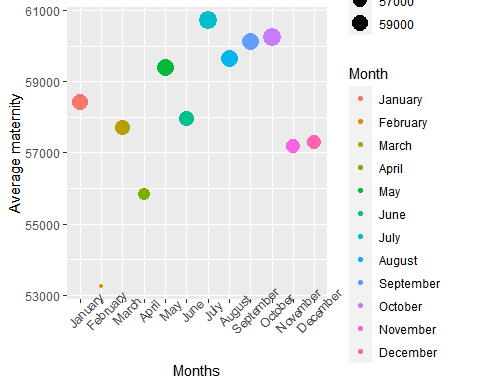
library("tidyverse")  
library("ggplot2")  
library("psych")  
library("dplyr")  
library("readr")  
library("tidyr")  
  
averages<- read\_csv("datasets/average admissions.csv")

## Rows: 12 Columns: 3

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (1): Month  
## dbl (2): Admissions, average maternity

##   
## 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.

averages$Month<-factor(averages$Month, levels = averages$Month)  
  
ggplot(averages, aes(Month, `average maternity`)) + geom\_point((aes(col=Month, size=`average maternity`))) + theme(axis.title.x = element\_text(vjust=0.5)) + theme(axis.text.x = element\_text(angle = 45)) + ylab("Average maternity") +xlab("Months")



corr.test(averages$Admissions, averages$`average maternity`)

## Call:corr.test(x = averages$Admissions, y = averages$`average maternity`)  
## Correlation matrix   
## [1] 0.4  
## Sample Size   
## [1] 12  
## These are the unadjusted probability values.  
## The probability values adjusted for multiple tests are in the p.adj object.   
## [1] 0.19  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

cor(averages$Admissions, averages$`average maternity`)

## [1] 0.4031278

simple.fit = lm(Admissions~`average maternity`, data=averages)   
summary(simple.fit)

##   
## Call:  
## lm(formula = Admissions ~ `average maternity`, data = averages)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -680291 -151794 35723 219372 497929   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.776e+06 2.891e+06 1.306 0.221  
## `average maternity` 6.920e+01 4.968e+01 1.393 0.194  
##   
## Residual standard error: 350900 on 10 degrees of freedom  
## Multiple R-squared: 0.1625, Adjusted R-squared: 0.07876   
## F-statistic: 1.94 on 1 and 10 DF, p-value: 0.1938

plot(simple.fit)

