Final\_ Project

Amy Richardson

4/6/2021

#Initiate packages

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.3

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.5 v dplyr 1.0.3  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 4.0.3

## Warning: package 'tibble' was built under R version 4.0.3

## Warning: package 'tidyr' was built under R version 4.0.3

## Warning: package 'readr' was built under R version 4.0.3

## Warning: package 'purrr' was built under R version 4.0.3

## Warning: package 'dplyr' was built under R version 4.0.3

## Warning: package 'stringr' was built under R version 4.0.3

## Warning: package 'forcats' was built under R version 4.0.3

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

library(ggplot2)  
library(cluster)  
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.0.4

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(dendextend)

## Warning: package 'dendextend' was built under R version 4.0.4

##   
## ---------------------  
## Welcome to dendextend version 1.14.0  
## Type citation('dendextend') for how to cite the package.  
##   
## Type browseVignettes(package = 'dendextend') for the package vignette.  
## The github page is: https://github.com/talgalili/dendextend/  
##   
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues  
## Or contact: <tal.galili@gmail.com>  
##   
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))  
## ---------------------

##   
## Attaching package: 'dendextend'

## The following object is masked from 'package:stats':  
##   
## cutree

library(dplyr)  
library(tidyr)  
library(psych)

## Warning: package 'psych' was built under R version 4.0.3

##   
## Attaching package: 'psych'

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

library(pastecs)

## Warning: package 'pastecs' was built under R version 4.0.4

##   
## Attaching package: 'pastecs'

## The following objects are masked from 'package:dplyr':  
##   
## first, last

## The following object is masked from 'package:tidyr':  
##   
## extract

library(car)

## Warning: package 'car' was built under R version 4.0.4

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.0.3

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

library(pgirmess)

## Warning: package 'pgirmess' was built under R version 4.0.4

##   
## Attaching package: 'pgirmess'

## The following object is masked from 'package:psych':  
##   
## shannon

##Looka at larger data set for VT students

Read in Data

degree\_all <- read.csv("VTNETSFinal.csv", header = TRUE)

Remove rows with NA For Transfer Institution

degree\_all<- filter(degree\_all, TransInst.Name != "NA")  
degree\_all <- filter(degree\_all, Eng.Bach.Flag == 1)  
degree\_all <- filter(degree\_all, NumberofCredits > 0)  
degree\_all <- select(degree\_all, TransInst.Name, NumberofCredits, Eng.Major, GPA, TotalSemesterTimetoDegree)

Create dataframe for Transient transfer students < 12 transfer credits

transient\_df <- filter(degree\_all, NumberofCredits < 12)  
transient\_df <- mutate(transient\_df, Xfer.Type = "Transient")

Create a dataframe with only VCCS transfers

vccs\_df <- filter(degree\_all, TransInst.Name == "New River Community College" | TransInst.Name == "Northern Va Cmty Coll-Annandal" | TransInst.Name == "Eastern Shore Community Coll"| TransInst.Name == "Virginia Western Cmty Coll" | TransInst.Name == "Tidewater Cmty Coll Nrflk-Main" | TransInst.Name == "Southwest Virginia Cmty Coll" | TransInst.Name == "J Sargeant Reynolds Cmty Coll" | TransInst.Name == "Piedmont Virginia Cmty College" | TransInst.Name == "Central Virginia Community Col" | TransInst.Name == "Thomas Nelson Community Coll" | TransInst.Name == "Germanna Community College" | TransInst.Name == "Danville Community College" | TransInst.Name == "Blue Ridge Cmty College VA" | TransInst.Name == "Patrick Henry Community Coll" | TransInst.Name =="Dabney Lancaster Cmty College" | TransInst.Name =="John Tyler Community College" | TransInst.Name == "Lord Fairfax Community Coll" | TransInst.Name == "Mountain Empire Cmty College" | TransInst.Name == "Rappahannock Community College" | TransInst.Name == "Southside VA Cmty Coll Alberta" | TransInst.Name == "Virginia Highlands Cmty Coll" | TransInst.Name =="Wytheville Community College" | TransInst.Name == "Northern Virginia Comm Coll")  
  
vccs\_df <- filter(vccs\_df, NumberofCredits > 12 )  
vccs\_df <- mutate(vccs\_df, Xfer.Type = "Vert.VCCS")

Create a dataframe with non-VCCS transfers

no\_vccs\_dfa <- anti\_join(degree\_all, vccs\_df)

## Joining, by = c("TransInst.Name", "NumberofCredits", "Eng.Major", "GPA", "TotalSemesterTimetoDegree")

no\_vccs\_dfa <- filter(no\_vccs\_dfa, TransInst.Name != "Credit by Exam")  
  
no\_vccs\_df <- filter(no\_vccs\_dfa, TransInst.Name == "Anne Arundel Community College"| TransInst.Name == "Austin Community Coll - TX"| TransInst.Name == "Barbados Comm Coll"| TransInst.Name == "Bergen Community College"| TransInst.Name == "Berkeley City College"| TransInst.Name == "Broward Cmty Coll Central"| TransInst.Name == "Carroll Cmty College"| TransInst.Name == "Chattanooga State Tech Cmty Cl"| TransInst.Name == "Clackamas Community College"| TransInst.Name == "Collin County Comm Coll" | TransInst.Name == "Colorado Northwestern Cmty Col"| TransInst.Name == "County College of Morris"| TransInst.Name == "Delaware County Community Coll" | TransInst.Name == "Delaware Tech CC Dover Terry C"| TransInst.Name == "Edmonds Community College"|TransInst.Name == "Everett Community College"| TransInst.Name == "Fiorello Laguardia Cmty Coll"| TransInst.Name == "Frederick Community College"| TransInst.Name == "Gateway Cmty College"| TransInst.Name == "Gateway Cmty Technical Coll"| TransInst.Name == "Green River Community College"| TransInst.Name == "Howard Community College"| TransInst.Name == "Hudson Valley Comm unity Coll"| TransInst.Name == "Mercer County Comm Coll" | TransInst.Name == "Miami-Dade Cmty Coll South"| TransInst.Name == "Miami-Dade College"| TransInst.Name == "Monroe Community College"| TransInst.Name == "Montgomery College" | TransInst.Name == "Montgomery College of Takoma" | TransInst.Name == "Montgomery College Rockville" | TransInst.Name == "Northampton Co Area Cmty Coll" | TransInst.Name == "Northampton Community College" | TransInst.Name == "Oakland Cmty Col Bloomfld" | TransInst.Name == "Onondaga Community College" | TransInst.Name == "Orange County Community Coll" | TransInst.Name == "Pasadena City College" | TransInst.Name == "Prince Georges Cmty College" | TransInst.Name == "Queensborough Community Coll" | TransInst.Name == "Raritan Valley Cmty Coll" | TransInst.Name == "Red Rocks Community College" | TransInst.Name == "Richard Bland College" | TransInst.Name == "Santa Fe Community College FL" | TransInst.Name == "Santa Fe College" | TransInst.Name == "San Diego City College" | TransInst.Name == "Seattle Central Community Coll" | TransInst.Name == "Spokane Falls Cmty College" | TransInst.Name == "Trident Tech College North" | TransInst.Name =="Westchester Community College")  
  
no\_vccs\_df <- filter(no\_vccs\_df, NumberofCredits > 12 )  
no\_vccs\_df <- mutate(no\_vccs\_df, Xfer.Type = "Vert")

Create a dataframe with Horizontal transfers

horz\_df <- anti\_join(no\_vccs\_dfa, no\_vccs\_df)

## Joining, by = c("TransInst.Name", "NumberofCredits", "Eng.Major", "GPA", "TotalSemesterTimetoDegree")

horz\_df <- filter(horz\_df, NumberofCredits > 12 )  
horz\_df <- mutate(horz\_df, Xfer.Type = "Horz")

Join all back to one dataframe

eng\_trans <- full\_join(transient\_df, vccs\_df)

## Joining, by = c("TransInst.Name", "NumberofCredits", "Eng.Major", "GPA", "TotalSemesterTimetoDegree", "Xfer.Type")

eng\_trans <- full\_join(eng\_trans, no\_vccs\_df)

## Joining, by = c("TransInst.Name", "NumberofCredits", "Eng.Major", "GPA", "TotalSemesterTimetoDegree", "Xfer.Type")

eng\_trans <- full\_join(eng\_trans, horz\_df)

## Joining, by = c("TransInst.Name", "NumberofCredits", "Eng.Major", "GPA", "TotalSemesterTimetoDegree", "Xfer.Type")

Combine the VCCS/non-VCCS data set for a Vertical Data Set

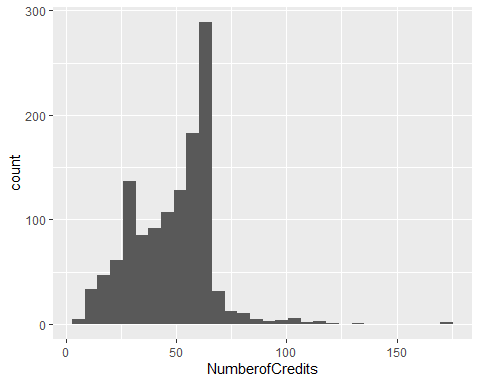
vert\_df <- full\_join(no\_vccs\_df, vccs\_df)

## Joining, by = c("TransInst.Name", "NumberofCredits", "Eng.Major", "GPA", "TotalSemesterTimetoDegree", "Xfer.Type")

##Explore Number of Credits

eng\_trans %>%  
 ggplot(aes(x = NumberofCredits)) +  
 geom\_histogram ()

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

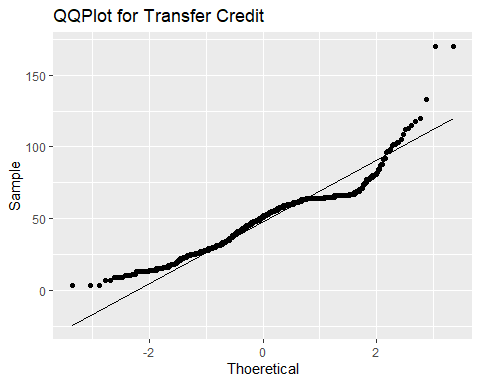


Q-Q Plot of Transfer Credit

qqplot.tsCredit <- qplot(sample = eng\_trans$NumberofCredits, stat="qq")

## Warning: `stat` is deprecated

qqplot.tsCredit + stat\_qq\_line()+  
 labs(title = "QQPlot for Transfer Credit", x = "Thoeretical", y = "Sample")



Shappiro-Wilks Test

shapiro.test(eng\_trans$NumberofCredits)

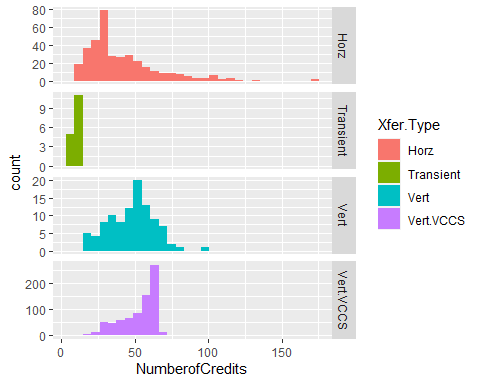
##   
## Shapiro-Wilk normality test  
##   
## data: eng\_trans$NumberofCredits  
## W = 0.94003, p-value < 2.2e-16

Since p < 0.05 the distribution of number of transfer credits is not normal

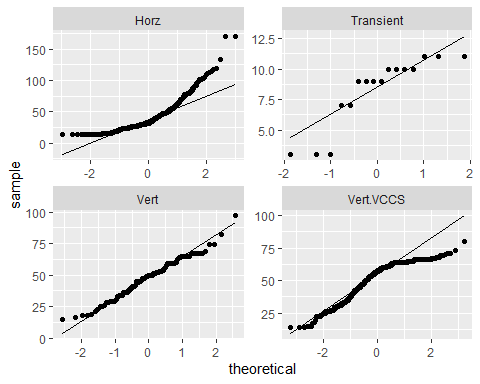
##Explore by Transfer Type Histograms by Engineering Discipline

eng\_trans %>%  
 ggplot(aes(x = NumberofCredits, fill = Xfer.Type)) +  
 geom\_histogram () +  
 facet\_grid(Xfer.Type ~., scales = "free")

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

 Q-Q plots by Transfer Type

eng\_trans %>%   
 ggplot(aes(sample=NumberofCredits)) +  
 stat\_qq() +  
 stat\_qq\_line() +  
 facet\_wrap(Xfer.Type ~ ., scales = "free")



Shappiro-Wilks Test Per Transfer Type

shapiro.test(horz\_df$NumberofCredits)

##   
## Shapiro-Wilk normality test  
##   
## data: horz\_df$NumberofCredits  
## W = 0.84502, p-value < 2.2e-16

shapiro.test(transient\_df$NumberofCredits)

##   
## Shapiro-Wilk normality test  
##   
## data: transient\_df$NumberofCredits  
## W = 0.79171, p-value = 0.002118

shapiro.test(no\_vccs\_df$NumberofCredits)

##   
## Shapiro-Wilk normality test  
##   
## data: no\_vccs\_df$NumberofCredits  
## W = 0.98278, p-value = 0.2174

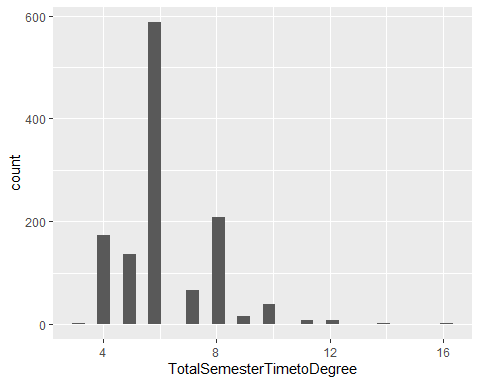
shapiro.test(vccs\_df$NumberofCredits)

##   
## Shapiro-Wilk normality test  
##   
## data: vccs\_df$NumberofCredits  
## W = 0.89433, p-value < 2.2e-16

## Checking Number of Semesters normality

eng\_trans %>%  
 ggplot(aes(x = TotalSemesterTimetoDegree)) +  
 geom\_histogram ()

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



Shappiro-Wilks Test

shapiro.test(eng\_trans$TotalSemesterTimetoDegree)

##   
## Shapiro-Wilk normality test  
##   
## data: eng\_trans$TotalSemesterTimetoDegree  
## W = 0.86819, p-value < 2.2e-16

# RQ1.How does the number of credits transferred vary across transfer groups?

##Kruskal=Wallis Test

kruskal.test(NumberofCredits ~ Xfer.Type, data = eng\_trans)

##   
## Kruskal-Wallis rank sum test  
##   
## data: NumberofCredits by Xfer.Type  
## Kruskal-Wallis chi-squared = 202.18, df = 3, p-value < 2.2e-16

eng\_trans$Ranks<-rank(eng\_trans$NumberofCredits)  
by(eng\_trans$Ranks, eng\_trans$Xfer.Type, mean)

## eng\_trans$Xfer.Type: Horz  
## [1] 439.5208  
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Transient  
## [1] 8.5  
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert  
## [1] 608.305  
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert.VCCS  
## [1] 724.9468

Because p < 0.05 we can conclude that Engineering Major does significantly impact the number of semesters to degree.

Post Hoc tests

kruskalmc(NumberofCredits ~ Xfer.Type, data = eng\_trans)

## Multiple comparison test after Kruskal-Wallis   
## p.value: 0.05   
## Comparisons  
## obs.dif critical.dif difference  
## Horz-Transient 431.0208 242.74468 TRUE  
## Horz-Vert 168.7842 107.39763 TRUE  
## Horz-Vert.VCCS 285.4260 60.64848 TRUE  
## Transient-Vert 599.8050 255.82091 TRUE  
## Transient-Vert.VCCS 716.4468 239.97568 TRUE  
## Vert-Vert.VCCS 116.6418 100.98315 TRUE

##Descriptive Statistics

describe(eng\_trans$NumberofCredits)

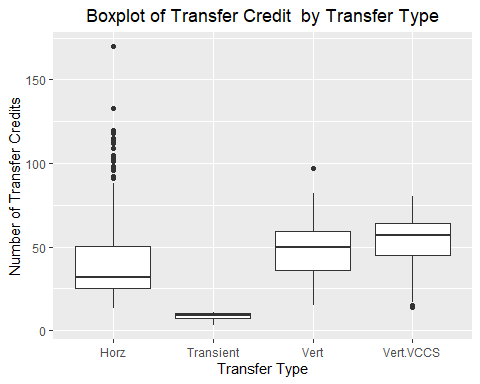
## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 1247 48.53 18.74 51 48.84 19.27 3 170 167 0.49 2.93 0.53

by(eng\_trans$NumberofCredits, eng\_trans$Xfer.Type, stat.desc)

## eng\_trans$Xfer.Type: Horz  
## nbr.val nbr.null nbr.na min max range   
## 3.600000e+02 0.000000e+00 0.000000e+00 1.300000e+01 1.700000e+02 1.570000e+02   
## sum median mean SE.mean CI.mean.0.95 var   
## 1.486750e+04 3.200000e+01 4.129861e+01 1.334600e+00 2.624616e+00 6.412163e+02   
## std.dev coef.var   
## 2.532225e+01 6.131501e-01   
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Transient  
## nbr.val nbr.null nbr.na min max range   
## 16.0000000 0.0000000 0.0000000 3.0000000 11.0000000 8.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 132.0000000 9.0000000 8.2500000 0.7158911 1.5258857 8.2000000   
## std.dev coef.var   
## 2.8635642 0.3470987   
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert  
## nbr.val nbr.null nbr.na min max range   
## 100.0000000 0.0000000 0.0000000 15.0000000 97.0000000 82.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 4793.5000000 50.0000000 47.9350000 1.5633694 3.1020641 244.4123990   
## std.dev coef.var   
## 15.6336943 0.3261436   
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert.VCCS  
## nbr.val nbr.null nbr.na min max range   
## 7.710000e+02 0.000000e+00 0.000000e+00 1.400000e+01 8.000000e+01 6.600000e+01   
## sum median mean SE.mean CI.mean.0.95 var   
## 4.072000e+04 5.700000e+01 5.281453e+01 4.572231e-01 8.975517e-01 1.611798e+02   
## std.dev coef.var   
## 1.269566e+01 2.403820e-01

Plot boxplots of transfer credits by transfer Type

eng\_trans %>%   
 ggplot(aes(x = Xfer.Type, y = NumberofCredits)) +  
 geom\_boxplot() +  
 labs(title = "Boxplot of Transfer Credit by Transfer Type",  
 x ="Transfer Type",  
 y = "Number of Transfer Credits") +  
 theme(plot.title = element\_text(hjust = 0.5))

 ## Difference between Vertical VCCS and non-VCCS

Exploratory analysis

by(vert\_df$NumberofCredits, vert\_df$Xfer.Type, stat.desc, basic=FALSE, norm=TRUE)

## vert\_df$Xfer.Type: Vert  
## median mean SE.mean CI.mean.0.95 var std.dev   
## 50.00000000 47.93500000 1.56336943 3.10206413 244.41239899 15.63369435   
## coef.var skewness skew.2SE kurtosis kurt.2SE normtest.W   
## 0.32614362 0.02833902 0.05870214 -0.08891330 -0.09294116 0.98277767   
## normtest.p   
## 0.21742352   
## ------------------------------------------------------------   
## vert\_df$Xfer.Type: Vert.VCCS  
## median mean SE.mean CI.mean.0.95 var   
## 5.700000e+01 5.281453e+01 4.572231e-01 8.975517e-01 1.611798e+02   
## std.dev coef.var skewness skew.2SE kurtosis   
## 1.269566e+01 2.403820e-01 -9.101798e-01 -5.168819e+00 -1.088194e-01   
## kurt.2SE normtest.W normtest.p   
## -3.093841e-01 8.943255e-01 1.393463e-22

leveneTest(vert\_df$NumberofCredits, vert\_df$Xfer.Type, center = "mean")

## Warning in leveneTest.default(vert\_df$NumberofCredits, vert\_df$Xfer.Type, :  
## vert\_df$Xfer.Type coerced to factor.

## Levene's Test for Homogeneity of Variance (center = "mean")  
## Df F value Pr(>F)   
## group 1 6.0323 0.01424 \*  
## 869   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

??Since Levene’s tests for Sunday and for Wednesday are not significantly different, the homogeneity of variance has been met.

Run Wilcoxon Rank-Sum TEst

vertModel<-wilcox.test(NumberofCredits ~ Xfer.Type, data = vert\_df, exact = FALSE, correct= FALSE)  
vertModel

##   
## Wilcoxon rank sum test  
##   
## data: NumberofCredits by Xfer.Type  
## W = 30474, p-value = 0.0006342  
## alternative hypothesis: true location shift is not equal to 0

Effect Size

rFromWilcox<-function(wilcoxModel, N){  
 z<- qnorm(wilcoxModel$p.value/2)  
 r<- z/ sqrt(N)  
 cat(wilcoxModel$data.name, "Effect Size, r = ", r)  
}  
  
rFromWilcox(vertModel, 20)

## NumberofCredits by Xfer.Type Effect Size, r = -0.7639611

# RQ2. How does the time to degree vary across transfer groups?

##Kruskal=Wallis Test

kruskal.test(TotalSemesterTimetoDegree ~ Xfer.Type, data = eng\_trans)

##   
## Kruskal-Wallis rank sum test  
##   
## data: TotalSemesterTimetoDegree by Xfer.Type  
## Kruskal-Wallis chi-squared = 28.863, df = 3, p-value = 2.393e-06

eng\_trans$Ranks<-rank(eng\_trans$TotalSemesterTimetoDegree)  
by(eng\_trans$Ranks, eng\_trans$Xfer.Type, mean)

## eng\_trans$Xfer.Type: Horz  
## [1] 677.7236  
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Transient  
## [1] 802.9688  
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert  
## [1] 701.565  
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert.VCCS  
## [1] 585.1407

Because p < 0.05 we can conclude that Engineering Major does significantly impact the number of semesters to degree.

Post Hoc tests

kruskalmc(TotalSemesterTimetoDegree ~ Xfer.Type, data = eng\_trans)

## Multiple comparison test after Kruskal-Wallis   
## p.value: 0.05   
## Comparisons  
## obs.dif critical.dif difference  
## Horz-Transient 125.24514 242.74468 FALSE  
## Horz-Vert 23.84139 107.39763 FALSE  
## Horz-Vert.VCCS 92.58288 60.64848 TRUE  
## Transient-Vert 101.40375 255.82091 FALSE  
## Transient-Vert.VCCS 217.82802 239.97568 FALSE  
## Vert-Vert.VCCS 116.42427 100.98315 TRUE

##Descriptive Statistics

describe(eng\_trans$TotalSemesterTimetoDegree)

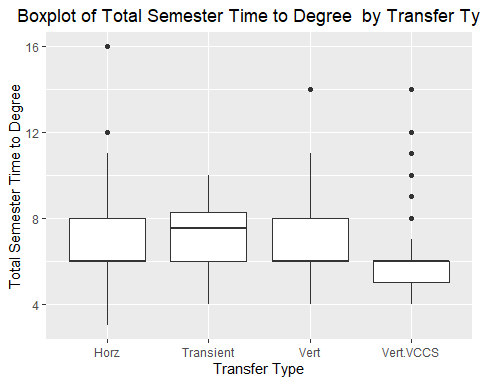
## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 1247 6.25 1.59 6 6.14 1.48 3 16 13 1.08 2.61 0.04

by(eng\_trans$TotalSemesterTimetoDegree, eng\_trans$Xfer.Type, stat.desc)

## eng\_trans$Xfer.Type: Horz  
## nbr.val nbr.null nbr.na min max range   
## 3.600000e+02 0.000000e+00 0.000000e+00 3.000000e+00 1.600000e+01 1.300000e+01   
## sum median mean SE.mean CI.mean.0.95 var   
## 2.315000e+03 6.000000e+00 6.430556e+00 8.091973e-02 1.591363e-01 2.357281e+00   
## std.dev coef.var   
## 1.535344e+00 2.387576e-01   
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Transient  
## nbr.val nbr.null nbr.na min max range   
## 16.0000000 0.0000000 0.0000000 4.0000000 10.0000000 6.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 114.0000000 7.5000000 7.1250000 0.4819665 1.0272872 3.7166667   
## std.dev coef.var   
## 1.9278658 0.2705777   
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert  
## nbr.val nbr.null nbr.na min max range   
## 100.0000000 0.0000000 0.0000000 4.0000000 14.0000000 10.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 659.0000000 6.0000000 6.5900000 0.1596049 0.3166908 2.5473737   
## std.dev coef.var   
## 1.5960494 0.2421926   
## ------------------------------------------------------------   
## eng\_trans$Xfer.Type: Vert.VCCS  
## nbr.val nbr.null nbr.na min max range   
## 7.710000e+02 0.000000e+00 0.000000e+00 4.000000e+00 1.400000e+01 1.000000e+01   
## sum median mean SE.mean CI.mean.0.95 var   
## 4.700000e+03 6.000000e+00 6.095979e+00 5.697138e-02 1.118377e-01 2.502464e+00   
## std.dev coef.var   
## 1.581918e+00 2.595019e-01

Plot boxplots of transfer credits by transfer Type

eng\_trans %>%   
 ggplot(aes(x = Xfer.Type, y = TotalSemesterTimetoDegree)) +  
 geom\_boxplot() +  
 labs(title = "Boxplot of Total Semester Time to Degree by Transfer Type",  
 x ="Transfer Type",  
 y = "Total Semester Time to Degree") +  
 theme(plot.title = element\_text(hjust = 0.5))



# RQ3. How does degree of engineering discipline relate to the number of semesters to degree after transfer for vertical transfer students?

Descriptive Statistics by Major for Vertical Transfer Students

describe(vert\_df$NumberofCredits)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 871 52.25 13.15 56 53.68 11.86 14 97 83 -0.78 -0.12 0.45

by(vert\_df$NumberofCredits, vert\_df$Eng.Major, stat.desc)

## vert\_df$Eng.Major: AE  
## nbr.val nbr.null nbr.na min max range   
## 58.000000 0.000000 0.000000 22.000000 82.000000 60.000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 2975.000000 50.000000 51.293103 1.662549 3.329198 160.316092   
## std.dev coef.var   
## 12.661599 0.246848   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: BSE  
## nbr.val nbr.null nbr.na min max range   
## 9.0000000 0.0000000 0.0000000 29.0000000 66.0000000 37.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 482.0000000 61.0000000 53.5555556 5.1937760 11.9768688 242.7777778   
## std.dev coef.var   
## 15.5813279 0.2909377   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: CE  
## nbr.val nbr.null nbr.na min max range   
## 156.0000000 0.0000000 0.0000000 16.0000000 74.0000000 58.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 8127.5000000 56.0000000 52.0993590 1.0333102 2.0411876 166.5658706   
## std.dev coef.var   
## 12.9060401 0.2477197   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: CHE  
## nbr.val nbr.null nbr.na min max range   
## 48.0000000 0.0000000 0.0000000 23.0000000 66.0000000 43.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 2354.0000000 55.0000000 49.0416667 1.9261541 3.8749221 178.0833333   
## std.dev coef.var   
## 13.3447867 0.2721112   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: CPE  
## nbr.val nbr.null nbr.na min max range   
## 61.0000000 0.0000000 0.0000000 17.0000000 69.0000000 52.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 3088.0000000 55.0000000 50.6229508 1.7827596 3.5660501 193.8721311   
## std.dev coef.var   
## 13.9237973 0.2750491   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: CS  
## nbr.val nbr.null nbr.na min max range   
## 84.0000000 0.0000000 0.0000000 15.0000000 74.0000000 59.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 4344.0000000 57.0000000 51.7142857 1.4565392 2.8969979 178.2065404   
## std.dev coef.var   
## 13.3494023 0.2581376   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: DAE  
## nbr.val nbr.null nbr.na min max range   
## 3.000000 0.000000 0.000000 33.000000 52.000000 19.000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 127.000000 42.000000 42.333333 5.487359 23.610201 90.333333   
## std.dev coef.var   
## 9.504385 0.224513   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: DOE  
## nbr.val nbr.null nbr.na min max range   
## 3.000000 0.000000 0.000000 33.000000 52.000000 19.000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 127.000000 42.000000 42.333333 5.487359 23.610201 90.333333   
## std.dev coef.var   
## 9.504385 0.224513   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: EE  
## nbr.val nbr.null nbr.na min max range   
## 108.0000000 0.0000000 0.0000000 14.0000000 68.0000000 54.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 5826.0000000 59.0000000 53.9444444 1.2373159 2.4528345 165.3426791   
## std.dev coef.var   
## 12.8585644 0.2383668   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: ESM  
## nbr.val nbr.null nbr.na min max range   
## 17.0000000 0.0000000 0.0000000 34.0000000 71.0000000 37.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 951.0000000 60.0000000 55.9411765 2.6505696 5.6189565 119.4338235   
## std.dev coef.var   
## 10.9285783 0.1953584   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: ISE  
## nbr.val nbr.null nbr.na min max range   
## 32.0000000 0.0000000 0.0000000 22.0000000 73.0000000 51.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 1611.0000000 52.0000000 50.3437500 2.3821452 4.8584171 181.5877016   
## std.dev coef.var   
## 13.4754481 0.2676687   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: ME  
## nbr.val nbr.null nbr.na min max range   
## 2.590000e+02 0.000000e+00 0.000000e+00 1.400000e+01 9.700000e+01 8.300000e+01   
## sum median mean SE.mean CI.mean.0.95 var   
## 1.379900e+04 5.800000e+01 5.327799e+01 8.206465e-01 1.616018e+00 1.744263e+02   
## std.dev coef.var   
## 1.320705e+01 2.478895e-01   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: MINE  
## nbr.val nbr.null nbr.na min max range   
## 9.0000000 0.0000000 0.0000000 39.0000000 67.0000000 28.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 514.0000000 64.0000000 57.1111111 3.4336749 7.9180686 106.1111111   
## std.dev coef.var   
## 10.3010248 0.1803681   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: MSE  
## nbr.val nbr.null nbr.na min max range   
## 22.0000000 0.0000000 0.0000000 26.0000000 66.0000000 40.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 1070.0000000 52.0000000 48.6363636 3.0730961 6.3908532 207.7662338   
## std.dev coef.var   
## 14.4140984 0.2963646   
## ------------------------------------------------------------   
## vert\_df$Eng.Major: OE  
## nbr.val nbr.null nbr.na min max range   
## 2.0000000 0.0000000 0.0000000 54.0000000 64.0000000 10.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 118.0000000 59.0000000 59.0000000 5.0000000 63.5310237 50.0000000   
## std.dev coef.var   
## 7.0710678 0.1198486

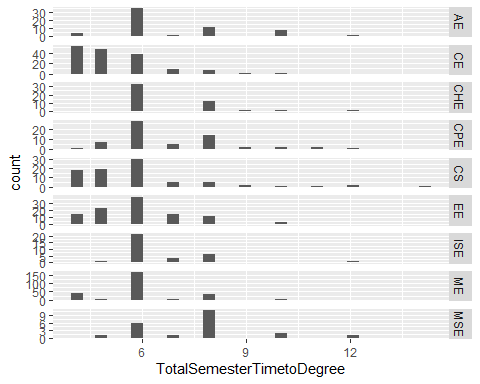
Remove Majors with less than 20 students

vert\_df\_sub <- filter(vert\_df, Eng.Major != "DAE")  
vert\_df\_sub <- filter(vert\_df\_sub, Eng.Major != "BSE")  
vert\_df\_sub<- filter(vert\_df\_sub, Eng.Major != "DOE")  
vert\_df\_sub <- filter(vert\_df\_sub, Eng.Major != "MINE")  
vert\_df\_sub <- filter(vert\_df\_sub, Eng.Major != "ESM")  
vert\_df\_sub<- filter(vert\_df\_sub, Eng.Major != "OE")

Histogram by Engineering Discipline

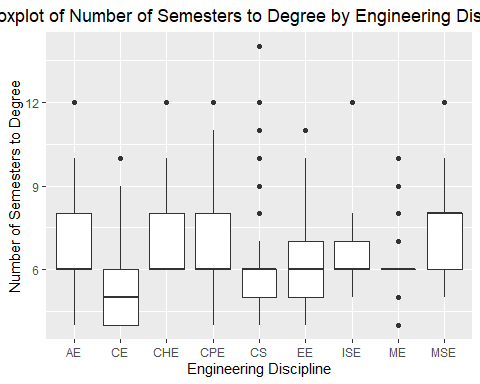
vert\_df\_sub %>%  
 ggplot(aes(x = TotalSemesterTimetoDegree)) +  
 geom\_histogram () +  
 facet\_grid(Eng.Major ~., scales = "free")

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



Plot boxplots of Total Semester to Degree by engineering discipline

vert\_df\_sub %>%   
 ggplot(aes(x = Eng.Major, y = TotalSemesterTimetoDegree)) +  
 geom\_boxplot() +  
 labs(title = "Boxplot of Number of Semesters to Degree by Engineering Discipline",  
 x ="Engineering Discipline",  
 y = "Number of Semesters to Degree") +  
 theme(plot.title = element\_text(hjust = 0.5))

 ##Kruskal=Wallis Test

kruskal.test(TotalSemesterTimetoDegree ~ Eng.Major, data = vert\_df\_sub)

##   
## Kruskal-Wallis rank sum test  
##   
## data: TotalSemesterTimetoDegree by Eng.Major  
## Kruskal-Wallis chi-squared = 146.13, df = 8, p-value < 2.2e-16

vert\_df\_sub$Ranks<-rank(vert\_df\_sub$TotalSemesterTimetoDegree)  
by(vert\_df\_sub$Ranks, vert\_df\_sub$Eng.Major, mean)

## vert\_df\_sub$Eng.Major: AE  
## [1] 525.0431  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: CE  
## [1] 261  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: CHE  
## [1] 529.7188  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: CPE  
## [1] 526.5984  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: CS  
## [1] 359.8393  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: EE  
## [1] 409.6528  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: ISE  
## [1] 512.7656  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: ME  
## [1] 423.7066  
## ------------------------------------------------------------   
## vert\_df\_sub$Eng.Major: MSE  
## [1] 630.5

Because p < 0.05 we can conclude that Engineering Major does significantly impact the number of semesters to degree.

Post Hoc tests

kruskalmc(TotalSemesterTimetoDegree ~ Eng.Major, data = vert\_df\_sub)

## Multiple comparison test after Kruskal-Wallis   
## p.value: 0.05   
## Comparisons  
## obs.dif critical.dif difference  
## AE-CE 264.043103 117.58943 TRUE  
## AE-CHE 4.675647 149.19560 FALSE  
## AE-CPE 1.555257 140.22720 FALSE  
## AE-CS 165.203818 130.53537 TRUE  
## AE-EE 115.390326 124.47033 FALSE  
## AE-ISE 12.277478 168.37203 FALSE  
## AE-ME 101.336540 111.07170 FALSE  
## AE-MSE 105.456897 191.45087 FALSE  
## CE-CHE 268.718750 126.20305 TRUE  
## CE-CPE 265.598361 115.46234 TRUE  
## CE-CS 98.839286 103.47641 FALSE  
## CE-EE 148.652778 95.71180 TRUE  
## CE-ISE 251.765625 148.38135 TRUE  
## CE-ME 162.706564 77.49069 TRUE  
## CE-MSE 369.500000 174.13018 TRUE  
## CHE-CPE 3.120389 147.52492 FALSE  
## CHE-CS 169.879464 138.34529 TRUE  
## CHE-EE 120.065972 132.63785 FALSE  
## CHE-ISE 16.953125 174.49664 FALSE  
## CHE-ME 106.012186 120.15349 FALSE  
## CHE-MSE 100.781250 196.85876 FALSE  
## CPE-CS 166.759075 128.62255 TRUE  
## CPE-EE 116.945583 122.46282 FALSE  
## CPE-ISE 13.832736 166.89343 FALSE  
## CPE-ME 102.891797 108.81728 FALSE  
## CPE-MSE 103.901639 190.15181 FALSE  
## CS-EE 49.813492 111.23380 FALSE  
## CS-ISE 152.926339 158.83714 FALSE  
## CS-ME 63.867278 96.00530 FALSE  
## CS-MSE 270.660714 183.12163 TRUE  
## EE-ISE 103.112847 153.89157 FALSE  
## EE-ME 14.053786 87.58081 FALSE  
## EE-MSE 220.847222 178.84885 TRUE  
## ISE-ME 89.059061 143.27134 FALSE  
## ISE-MSE 117.734375 211.76210 FALSE  
## ME-MSE 206.793436 169.79685 TRUE

# RQ4. The median semester to degrees 4 for vertical transfer students is 4.

# RQ5. ##Correlation Analaysis between Transfer Credits and Number of semesters to degree for each Transfer Type

## Transient

Kendall’s Tau

cor.test(transient\_df$NumberofCredits, transient\_df$TotalSemesterTimetoDegree, alternative = "less", method="kendall")

## Warning in cor.test.default(transient\_df$NumberofCredits,  
## transient\_df$TotalSemesterTimetoDegree, : Cannot compute exact p-value with ties

##   
## Kendall's rank correlation tau  
##   
## data: transient\_df$NumberofCredits and transient\_df$TotalSemesterTimetoDegree  
## z = 0.75441, p-value = 0.7747  
## alternative hypothesis: true tau is less than 0  
## sample estimates:  
## tau   
## 0.1553691

Spearman’s Correlation

cor.test(transient\_df$NumberofCredits, transient\_df$TotalSemesterTimetoDegree, alternative = "less", method="spearman")

## Warning in cor.test.default(transient\_df$NumberofCredits,  
## transient\_df$TotalSemesterTimetoDegree, : Cannot compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: transient\_df$NumberofCredits and transient\_df$TotalSemesterTimetoDegree  
## S = 586.94, p-value = 0.6934  
## alternative hypothesis: true rho is less than 0  
## sample estimates:  
## rho   
## 0.1368516

##Horizontal Kendall’s Tau

cor.test(horz\_df$NumberofCredits, horz\_df$TotalSemesterTimetoDegree, alternative = "less", method="kendall")

##   
## Kendall's rank correlation tau  
##   
## data: horz\_df$NumberofCredits and horz\_df$TotalSemesterTimetoDegree  
## z = -9.1021, p-value < 2.2e-16  
## alternative hypothesis: true tau is less than 0  
## sample estimates:  
## tau   
## -0.3633109

Spearman’s Correlation

cor.test(horz\_df$NumberofCredits, horz\_df$TotalSemesterTimetoDegree, alternative = "less", method="spearman")

## Warning in cor.test.default(horz\_df$NumberofCredits,  
## horz\_df$TotalSemesterTimetoDegree, : Cannot compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: horz\_df$NumberofCredits and horz\_df$TotalSemesterTimetoDegree  
## S = 11413941, p-value < 2.2e-16  
## alternative hypothesis: true rho is less than 0  
## sample estimates:  
## rho   
## -0.4678536

##non-VCCS Kendall’s Tau

cor.test(no\_vccs\_df$NumberofCredits, no\_vccs\_df$TotalSemesterTimetoDegree, alternative = "less", method="kendall")

##   
## Kendall's rank correlation tau  
##   
## data: no\_vccs\_df$NumberofCredits and no\_vccs\_df$TotalSemesterTimetoDegree  
## z = -0.86807, p-value = 0.1927  
## alternative hypothesis: true tau is less than 0  
## sample estimates:  
## tau   
## -0.06751105

Spearman’s Correlation

cor.test(no\_vccs\_df$NumberofCredits, no\_vccs\_df$TotalSemesterTimetoDegree, alternative = "less", method="spearman")

## Warning in cor.test.default(no\_vccs\_df$NumberofCredits,  
## no\_vccs\_df$TotalSemesterTimetoDegree, : Cannot compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: no\_vccs\_df$NumberofCredits and no\_vccs\_df$TotalSemesterTimetoDegree  
## S = 181205, p-value = 0.1938  
## alternative hypothesis: true rho is less than 0  
## sample estimates:  
## rho   
## -0.08733693

##VCCS Kendall’s Tau

cor.test(vccs\_df$NumberofCredits, vccs\_df$TotalSemesterTimetoDegree, alternative = "less", method="kendall")

##   
## Kendall's rank correlation tau  
##   
## data: vccs\_df$NumberofCredits and vccs\_df$TotalSemesterTimetoDegree  
## z = -7.2882, p-value = 1.571e-13  
## alternative hypothesis: true tau is less than 0  
## sample estimates:  
## tau   
## -0.1985257

Spearman’s Correlation

cor.test(vccs\_df$NumberofCredits, vccs\_df$TotalSemesterTimetoDegree, alternative = "less", method="spearman")

## Warning in cor.test.default(vccs\_df$NumberofCredits,  
## vccs\_df$TotalSemesterTimetoDegree, : Cannot compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: vccs\_df$NumberofCredits and vccs\_df$TotalSemesterTimetoDegree  
## S = 96391449, p-value = 7.346e-14  
## alternative hypothesis: true rho is less than 0  
## sample estimates:  
## rho   
## -0.261907

##END