FINM4411

Most of the early code here is not pertinent and just used for other regressions and to load stuff. Skip to **SKIP TO HERE** for more pertinent stuff.

Java Installation (used for Powerpoint interface)

```
require(rJava)

## Loading required package: rJava

system("java -version")
```

Package Installation, not pertinent

```
#install.packages('quantmod')
#install.packages('xts')
#install.packages("ggplot2")
#install.packages("data.table")
#install.packages("gridExtra")
#install.packages("knitr")
#install.packages("zoo")
#install.packages("psych")
#install.packages("fBasics") ## online download, needed to find summary
statistics
#install.packages("kableExtra") ## online download, needed for Kable function
#install.packages("tidyverse") ## online download, needed for pull() function
#install.packages("lubridate")
#install.packages("lfe")
#install.packages("reshape2")
#install.packages("boot")
#install.packages("dplyr")
#install.packages("RCurl")
#install.packages("packrat")
#install.packages("rsconnect")
#install.packages("officer")
#install.packages("rvg")
#install.packages("scales")
#install.packages("magrittr")
#install.packages("devtools")
#install.packages("here")
#install.packages("glue")
#install.packages("viridis")
#install.packages("xlsx")
#install.packages("caret")
```

Library call, not pertinent

```
library(xts) ## 'r load-packages' command used above so Library is called
during Knit process
library(quantmod)
library(data.table)
library(ggplot2)
library(gridExtra)
library(fBasics)
library(knitr)
library(zoo)
library(psych)
library(kableExtra)
library(tidyverse)
library(lubridate)
library(lfe)
library(reshape2)
library(boot)
library(dplyr)
library(RCurl)
library(packrat)
library(rsconnect)
library(officer)
library(rvg)
library(scales)      # for formatting numbers
library(magrittr)    # for the %>% operator
library(devtools)
library(rJava)
library(customLayout)
library(here)
library(glue)
library(xlsx)
```

Preqin data download

```
Preqin1 <- read.csv('Preqin1 Values Only.csv', row.names = NULL,
stringsAsFactors = FALSE)
## above code saves Preqin1 Values Only.csv data into the variable 'Preqin1'
```

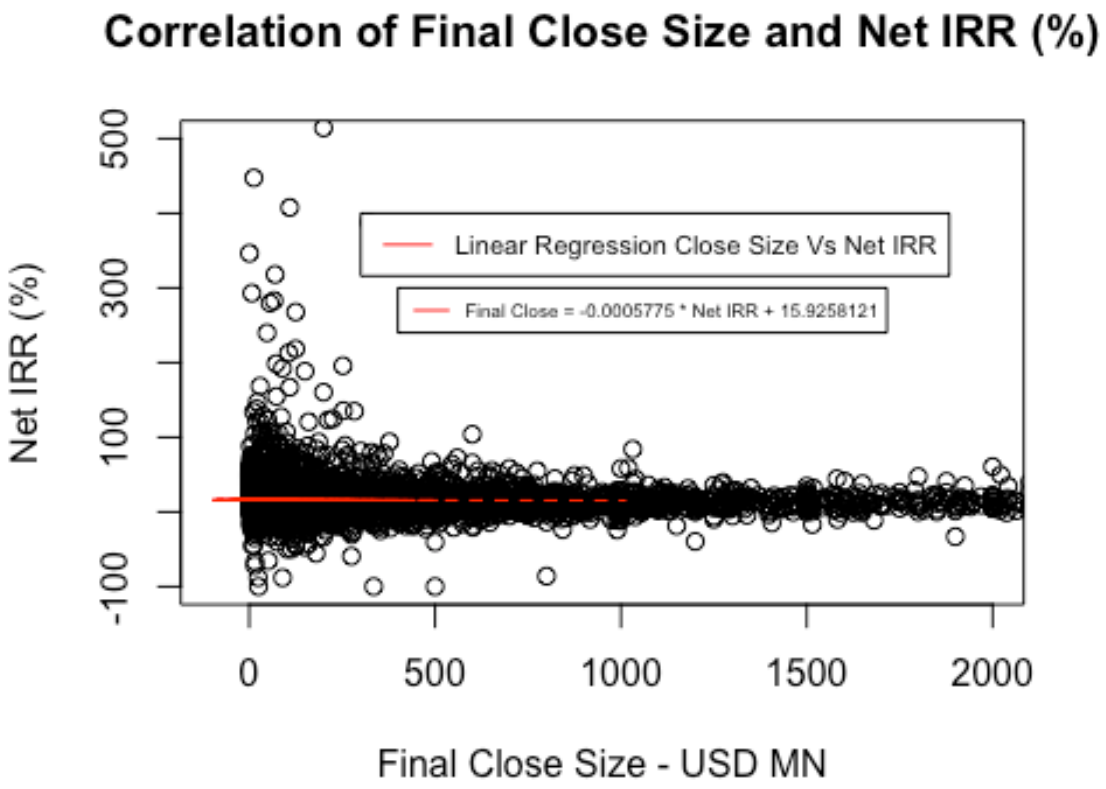
OLS regression of Final Close Size - US MN vs Net IRR

```
mod <- lm(Preqin1$NET.IRR.... ~ Preqin1$FINAL.CLOSE.SIZE..USD.MN., na.exclude
= TRUE)
## performs linear regression between Final Close Size - US MN vs Net IRR

summary(mod)

##
## Call:
## lm(formula = Preqin1$NET.IRR.... ~ Preqin1$FINAL.CLOSE.SIZE..USD.MN.,
##     na.exclude = TRUE)
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
## -115.9   -9.5   -3.3    4.9  997.2
##
## Coefficients:
##                                    Estimate Std. Error t value
## (Intercept)                       15.925812   0.408506   38.99
## Preqin1$FINAL.CLOSE.SIZE..USD.MN. -0.000578   0.000258   -2.24
##                                              Pr(>|t|)
## (Intercept)                       <0.0000000000000002 ***
## Preqin1$FINAL.CLOSE.SIZE..USD.MN.               0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.9 on 5532 degrees of freedom
## Multiple R-squared:  0.000904,   Adjusted R-squared:  0.000723
## F-statistic:    5 on 1 and 5532 DF,  p-value: 0.0253
```

Below code creates scatter plot of Fund Final Close Size vs IRR

```
plot(Preqin1$FINAL.CLOSE.SIZE..USD.MN., Preqin1$NET.IRR...., xlim=c(-
100,2000), ylim=c(-100,500), main="Correlation of Final Close Size and Net
IRR (%)", ylab = "Net IRR (%)", xlab = "Final Close Size - USD MN")
## above code plots points on the graph

## plotting fitted Security Market Line (SML) using regression coefficients
from Question 7.d
 lines(Preqin1$NET.IRR...., -0.0005775 * Preqin1$NET.IRR.... + 15.9258121,
col="red",lty=2)
 legend(300, 400, legend=c("Linear Regression Close Size Vs Net IRR"),
       col=c("red"), lty=1:2, cex=0.7)
 legend(400, 300, legend=c("Final Close = -0.0005775 * Net IRR +
15.9258121"),
       col=c("red"), lty=1:2, cex=0.5)
```



Preqin data download using different data set

```
Preqin2<- read.csv('Preqin2 Values Only.csv', row.names = NULL,
stringsAsFactors = FALSE)
## above code saves Preqin2 Values Only.csv data into the variable 'Preqin2'
```

Sorts data by IRR, below shows the top 6 performing funds in the data set by Net IRR. As shown most funds are quite small with $322 Million beng the largest. All but 2 are venture capital firms. Surprisingly, 3 out of the 6 have vintage years before 2000 which is interesting how they maintained such returns for so long. Another interesting thing is that all but 2 of them are currently liquidated. Also, all but 1 fund were/are located in the US.

```
TopIRR <- Preqin2[order(Preqin2$NET.IRR....),]
#tail(TopIRR) use this command to view the data but it doesn't Knit to pdf
well
```

Splitting data and for ease of use and finding basic mean IRR of US data

```
Geo<-split(Preqin2,Preqin2$GEOGRAPHIC.FOCUS)
USmean <- mean(Geo$US$NET.IRR....)
```

Creating dummy variables Vintage year using funds incepted before 2000 as Vin (Vintage)

```
Condense <- as.data.frame(cbind(Preqin2$VINTAGE...INCEPTION.YEAR,
Preqin2$NET.IRR....))
colnames(Condense) <- c("Date","IRR")
#creates data set with only Vintage and net IRR

for(i in 1:length(Condense$Date))
{
    if(Condense[i,"Date"]<2000 & Condense[i,"Date"]>1995)
    {
      Condense$Vin[i]=1
    }
    else
    {
      Condense$Vin[i]=0
    }
}
```

Creating dummy variables for Geographic Focus i.e. U.S. = 1, EUR = 0, ASIA = 0 for US data. Repeat process for different Geographic Focus areas.

```
Condense2 <- as.data.frame(cbind(Preqin2$GEOGRAPHIC.FOCUS,
Preqin2$NET.IRR....))
colnames(Condense2) <- c("GeoFocus","IRR")
#creates data set with only Geographic focus and net IRR data

#Below code creates US dummy variable
for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "US")
    {
      Condense2$US[i]=1
    }
    else
    {
      Condense2$US[i]=0
    }
}

#Below code creates Asia dummy variable
for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "Asia")
    {
      Condense2$Asia[i]=1
    }
    else
    {
      Condense2$Asia[i]=0
    }
}
```

```r
#Below code creates India dummy variable
for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "India")
    {
      Condense2$India[i]=1
    }
    else
    {
      Condense2$India[i]=0
    }
}

#Below code creates Australia dummy variable
for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "Australia")
    {
      Condense2$Australia[i]=1
    }
    else
    {
      Condense2$Australia[i]=0
    }
}

#Below code creates Africa dummy variable
for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "Africa")
    {
      Condense2$Africa[i]=1
    }
    else
    {
      Condense2$Africa[i]=0
    }
}

#Below code creates Europe dummy variable
for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "Europe")
    {
      Condense2$Europe[i]=1
    }
    else
    {
```

```
      Condense2$Europe[i]=0
    }
}

for(i in 1:length(Condense2$GeoFocus))
{
    if(Condense2[i,"GeoFocus"] == "Brazil")
    {
      Condense2$Brazil[i]=1
    }
    else
    {
      Condense2$Brazil[i]=0
    }
}
```

Multiple Regression

```
ols_no_int <- lm(Condense2$IRR ~ Condense2$US + Condense2$Asia +
Condense2$India + Condense2$Australia + Condense2$Africa + Condense2$Brazil -
1 )
# -1 means exclude intercept

summary(ols_no_int)

##
## Call:
## lm(formula = Condense2$IRR ~ Condense2$US + Condense2$Asia +
##     Condense2$India + Condense2$Australia + Condense2$Africa +
##     Condense2$Brazil - 1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -103.5   -2.6    8.1   16.9 1013.0
##
## Coefficients:
##                      Estimate Std. Error t value             Pr(>|t|)
## Condense2$US           15.587      0.729   21.39 < 0.0000000000000002 ***
## Condense2$Asia         13.047      1.890    6.90      0.0000000000057 ***
## Condense2$India        17.716      4.963    3.57              0.00036 ***
## Condense2$Australia    18.287      3.917    4.67      0.0000031082287 ***
## Condense2$Africa        9.635      7.019    1.37              0.16989
## Condense2$Brazil       13.530      6.676    2.03              0.04275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.6 on 5415 degrees of freedom
## Multiple R-squared:  0.0916, Adjusted R-squared:  0.0906
## F-statistic:   91 on 6 and 5415 DF,  p-value: <0.0000000000000002
```

T-Test to determine if there is a significant difference in the net IRR's of the U.S. and Australia

```
t.test(c(Geo$US$NET.IRR....),c(Geo$Australia$NET.IRR....))

##
##  Welch Two Sample t-test
##
## data:  c(Geo$US$NET.IRR....) and c(Geo$Australia$NET.IRR....)
## t = -1.171, df = 73.88, p-value = 0.246
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.29636  1.89632
## sample estimates:
## mean of x mean of y
##   15.5865   18.2866

## Performs Welch Two Sample t-test
```

Cleaned Preqin Data

```
Preqinclean<- read.csv('Calerecent.csv', row.names = NULL, stringsAsFactors =
FALSE)
## above code saves Calerecent.csv data into the variable 'Preqinclean'

Preqin_mod <- as.data.frame(cbind(Preqinclean$NET.IRR....,
Preqinclean$FUND.SIZE..USD.MN., Preqinclean$Market.rate...Y.N...,
Preqinclean$Impact...Y.N..., Preqinclean$STRATEGY,
Preqinclean$PRIMARY.REGION.FOCUS, Preqinclean$VINTAGE...INCEPTION.YEAR))
colnames(Preqin_mod) <- c("IRR","Fundsize","Market_Rate?","Impact?",
"Strategy","RegionFocus","Vintage")
#creates data set with only Market Rate and net IRR data
```

Creating Dummy variables

```
#Below code creates market return dummy variable
for(i in 1:length(Preqin_mod$`Market_Rate?`))
{
    if(Preqin_mod[i,"Market_Rate?"] == "Y")
    {
      Preqin_mod$MarketDUM[i]=1
    }
    else
    {
      Preqin_mod$MarketDUM[i]=0
    }
}

head(Preqin_mod)
```

```
##     IRR Fundsize Market_Rate? Impact?    Strategy   RegionFocus Vintage
## 1  7.88      167            Y       N Early Stage North America    2007
## 2 21.79      175            Y       N Early Stage North America    2011
## 3 11.51      217            Y       N Early Stage North America    2015
## 4  11.6    203.4            Y       N      Growth North America    2015
## 5  11.4      475            Y       N      Buyout North America    1993
## 6  -5.9      530            Y       N      Buyout North America    1998
##   MarketDUM
## 1         1
## 2         1
## 3         1
## 4         1
## 5         1
## 6         1

#Below code creates impact dummy variable where dummy = 1 for Y
for(i in 1:length(Preqinclean$Impact...Y.N...))
{
    if(Preqin_mod[i,"Impact?"] == "Y")
    {
      Preqin_mod$ImpactDUM[i]=1
    }
    else
    {
      Preqin_mod$ImpactDUM[i]=0
    }
}


#Creating more dummy variables for Strategy i.e. Early stage = 1, Growth = 0,
Buyout = 0 for US data. Repeat process for different categorical strategies.

#Below code strategy dummy variables
for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Early Stage")
    {
      Preqin_mod$Early_StageDUM[i]=1
    }
    else
    {
      Preqin_mod$Early_StageDUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Growth")
    {
      Preqin_mod$GrowthDUM[i]=1
```

```
    }
    else
    {
      Preqin_mod$GrowthDUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Buyout")
    {
      Preqin_mod$BuyoutDUM[i]=1
    }
    else
    {
      Preqin_mod$BuyoutDUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Fund of Funds")
    {
      Preqin_mod$Fund_of_FundsDUM[i]=1
    }
    else
    {
      Preqin_mod$Fund_of_FundsDUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Venture (General)")
    {
      Preqin_mod$Venture_GeneralDUM[i]=1
    }
    else
    {
      Preqin_mod$Venture_GeneralDUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Early Stage: Seed")
    {
      Preqin_mod$Early_Stage_SeedDUM[i]=1
    }
```

```
    else
    {
      Preqin_mod$Early_Stage_SeedDUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$Strategy))
{
    if(Preqin_mod[i,"Strategy"] == "Co-Investment")
    {
      Preqin_mod$Co_InvestmentDUM[i]=1
    }
    else
    {
      Preqin_mod$Co_InvestmentDUM[i]=0
    }
}

#Below code creates North America dummy variable
for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "North America")
    {
      Preqin_mod$North_America_DUM[i]=1
    }
    else
    {
      Preqin_mod$North_America_DUM[i]=0
    }
}

#Below code creates Europe dummy variable
for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "Europe")
    {
      Preqin_mod$Europe_DUM[i]=1
    }
    else
    {
      Preqin_mod$Europe_DUM[i]=0
    }
}

#Below code creates Asia dummy variable
for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "Asia")
    {
```

```
      Preqin_mod$Asia_DUM[i]=1
    }
    else
    {
      Preqin_mod$Asia_DUM[i]=0
    }
}

#Below code creates Diversified Multi-Regional dummy variable
for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "Diversified Multi-Regional")
    {
      Preqin_mod$Diversified_Multi_Regional_DUM[i]=1
    }
    else
    {
      Preqin_mod$Diversified_Multi_Regional_DUM[i]=0
    }
}

#Below code creates Americas dummy variable
for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "Americas")
    {
      Preqin_mod$Americas_DUM[i]=1
    }
    else
    {
      Preqin_mod$Americas_DUM[i]=0
    }
}

#Below code creates Africa dummy variable
for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "Africa")
    {
      Preqin_mod$Africa_DUM[i]=1
    }
    else
    {
      Preqin_mod$Africa_DUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$RegionFocus))
{
```

```
    if(Preqin_mod[i,"RegionFocus"] == "Middle East & Israel")
    {
      Preqin_mod$Middle_East_and_Israel_DUM[i]=1
    }
    else
    {
      Preqin_mod$Middle_East_and_Israel_DUM[i]=0
    }
}

for(i in 1:length(Preqin_mod$RegionFocus))
{
    if(Preqin_mod[i,"RegionFocus"] == "Australasia")
    {
      Preqin_mod$Australasia_DUM[i]=1
    }
    else
    {
      Preqin_mod$Australasia_DUM[i]=0
    }
}
```

Below code creates a matrix of dummy variables with 1969 - 2017 as columns and the length of the total dataset as rows.

```
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
k <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
n <- length(Preqinclean$NET.IRR....)


IRR <- matrix(NA, nrow=n, ncol=k)

for(j in 1:k){
for(i in 1:n)
{
    if(Preqinclean$VINTAGE...INCEPTION.YEAR[i] == sum((min-1)+j))
    {
      IRR[i,j]=1
    }
    else
    {
      IRR[i,j]=0
    }
}
}

Vintage_all <- IRR[,1:k]
colnames(Vintage_all) <- c(1969:2017)
Vintage <- Vintage_all[,c(1,3,4,8:49)]
# above code removes vintage years with no data
```

SKIP TO HERE Multiple Regression for Dummy variables

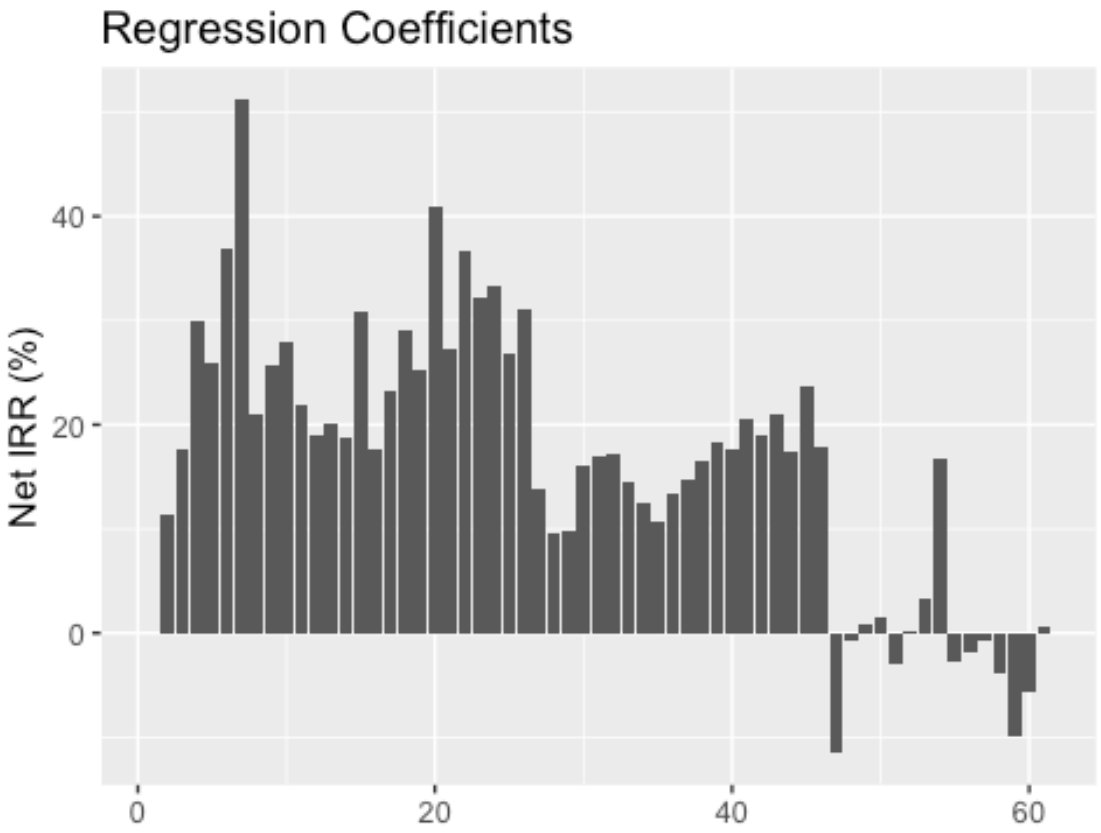
```
ols_no_int1 <- lm(Preqinclean$NET.IRR.... ~ Preqinclean$FUND.SIZE..USD.MN. +
Vintage  + Preqin_mod$ImpactDUM + Preqin_mod$Early_StageDUM +
Preqin_mod$GrowthDUM + Preqin_mod$BuyoutDUM + Preqin_mod$Fund_of_FundsDUM +
Preqin_mod$Venture_GeneralDUM + Preqin_mod$Early_Stage_SeedDUM +
Preqin_mod$Co_InvestmentDUM + Preqin_mod$North_America_DUM +
Preqin_mod$Europe_DUM + Preqin_mod$Asia_DUM +
Preqin_mod$Diversified_Multi_Regional_DUM + Preqin_mod$Americas_DUM +
Preqin_mod$Africa_DUM + Preqin_mod$Middle_East_and_Israel_DUM - 1)

summary(ols_no_int1)

##
## Call:
## lm(formula = Preqinclean$NET.IRR.... ~ Preqinclean$FUND.SIZE..USD.MN. +
##     Vintage + Preqin_mod$ImpactDUM + Preqin_mod$Early_StageDUM +
##     Preqin_mod$GrowthDUM + Preqin_mod$BuyoutDUM +
Preqin_mod$Fund_of_FundsDUM +
##     Preqin_mod$Venture_GeneralDUM + Preqin_mod$Early_Stage_SeedDUM +
##     Preqin_mod$Co_InvestmentDUM + Preqin_mod$North_America_DUM +
##     Preqin_mod$Europe_DUM + Preqin_mod$Asia_DUM +
Preqin_mod$Diversified_Multi_Regional_DUM +
##     Preqin_mod$Americas_DUM + Preqin_mod$Africa_DUM +
Preqin_mod$Middle_East_and_Israel_DUM -
##     1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -109.2   -9.4   -1.7    5.3  975.4
##
## Coefficients:
##                                            Estimate Std. Error t value
## Preqinclean$FUND.SIZE..USD.MN.            -0.000406   0.000269   -1.51
## Vintage1969                               11.284486  27.437140    0.41
## Vintage1971                               17.673024  27.402551    0.64
## Vintage1972                               29.883942  19.524361    1.53
## Vintage1976                               25.901488  27.432697    0.94
## Vintage1977                               36.789712  27.433862    1.34
## Vintage1978                               51.140781  19.556103    2.62
## Vintage1979                               20.894998  19.545099    1.07
## Vintage1980                               25.768722   9.674835    2.66
## Vintage1981                               28.042823  12.657122    2.22
## Vintage1982                               21.810299   9.658786    2.26
## Vintage1983                               18.920370   7.569921    2.50
## Vintage1984                               20.130086   6.503893    3.10
## Vintage1985                               18.802448   6.830149    2.75
## Vintage1986                               30.741341   6.223076    4.94
## Vintage1987                               17.673100   5.750311    3.07
## Vintage1988                               23.173859   5.568281    4.16
```

```
## Vintage1989                                  29.088538   5.335792    5.45
## Vintage1990                                  25.154619   5.225110    4.81
## Vintage1991                                  40.938513   6.194072    6.61
## Vintage1992                                  27.362590   4.819256    5.68
## Vintage1993                                  36.609729   4.650684    7.87
## Vintage1994                                  32.283172   4.518855    7.14
## Vintage1995                                  33.360303   4.391574    7.60
## Vintage1996                                  26.892273   4.351530    6.18
## Vintage1997                                  30.995553   4.039081    7.67
## Vintage1998                                  13.873730   3.826120    3.63
## Vintage1999                                   9.667120   3.885471    2.49
## Vintage2000                                   9.891320   3.724456    2.66
## Vintage2001                                  16.009512   3.847051    4.16
## Vintage2002                                  17.072216   4.069279    4.20
## Vintage2003                                  17.110711   4.080265    4.19
## Vintage2004                                  14.462197   3.865120    3.74
## Vintage2005                                  12.447448   3.675453    3.39
## Vintage2006                                  10.609798   3.630801    2.92
## Vintage2007                                  13.305622   3.587058    3.71
## Vintage2008                                  14.634103   3.611065    4.05
## Vintage2009                                  16.523450   3.996982    4.13
## Vintage2010                                  18.197432   3.849290    4.73
## Vintage2011                                  17.581732   3.721743    4.72
## Vintage2012                                  20.474217   3.732381    5.49
## Vintage2013                                  18.882504   3.682767    5.13
## Vintage2014                                  21.069317   3.631465    5.80
## Vintage2015                                  17.437361   3.629764    4.80
## Vintage2016                                  23.740167   3.596752    6.60
## Vintage2017                                  17.936597   3.725838    4.81
## Preqin_mod$ImpactDUM                        -11.433887   6.062649   -1.89
## Preqin_mod$Early_StageDUM                    -0.755503   1.679796   -0.45
## Preqin_mod$GrowthDUM                          0.825214   1.705758    0.48
## Preqin_mod$BuyoutDUM                          1.476570   1.213141    1.22
## Preqin_mod$Fund_of_FundsDUM                  -2.930747   1.351989   -2.17
## Preqin_mod$Venture_GeneralDUM                 0.173512   1.394699    0.12
## Preqin_mod$Early_Stage_SeedDUM                3.393560   3.218596    1.05
## Preqin_mod$Co_InvestmentDUM                  16.637105   3.147412    5.29
## Preqin_mod$North_America_DUM                 -2.756373   3.192758   -0.86
## Preqin_mod$Europe_DUM                        -1.929736   3.253272   -0.59
## Preqin_mod$Asia_DUM                          -0.681247   3.395362   -0.20
## Preqin_mod$Diversified_Multi_Regional_DUM    -3.839031   4.892482   -0.78
## Preqin_mod$Americas_DUM                      -9.965697   4.291868   -2.32
## Preqin_mod$Africa_DUM                        -5.742518   5.432838   -1.06
## Preqin_mod$Middle_East_and_Israel_DUM         0.596627   4.636518    0.13
##                                                      Pr(>|t|)
## Preqinclean$FUND.SIZE..USD.MN.                        0.13063
## Vintage1969                                           0.68088
## Vintage1971                                           0.51899
## Vintage1972                                           0.12593
## Vintage1976                                           0.34512
```

```
## Vintage1977                                    0.17997
## Vintage1978                                    0.00895 **
## Vintage1979                                    0.28509
## Vintage1980                                    0.00776 **
## Vintage1981                                    0.02676 *
## Vintage1982                                    0.02398 *
## Vintage1983                                    0.01247 *
## Vintage1984                                    0.00198 **
## Vintage1985                                    0.00593 **
## Vintage1986                          0.0000008050909960 ***
## Vintage1987                                    0.00213 **
## Vintage1988                          0.0000320686696775 ***
## Vintage1989                          0.0000000521273469 ***
## Vintage1990                          0.0000015181520002 ***
## Vintage1991                          0.0000000000423078 ***
## Vintage1992                          0.0000000143551236 ***
## Vintage1993                          0.0000000000000042 ***
## Vintage1994                          0.0000000000010253 ***
## Vintage1995                          0.0000000000000356 ***
## Vintage1996                          0.0000000006879094 ***
## Vintage1997                          0.0000000000000197 ***
## Vintage1998                                    0.00029 ***
## Vintage1999                                    0.01288 *
## Vintage2000                                    0.00794 **
## Vintage2001                          0.0000321049192452 ***
## Vintage2002                          0.0000276746150428 ***
## Vintage2003                          0.0000279020973037 ***
## Vintage2004                                    0.00018 ***
## Vintage2005                                    0.00071 ***
## Vintage2006                                    0.00349 **
## Vintage2007                                    0.00021 ***
## Vintage2008                          0.0000513654840943 ***
## Vintage2009                          0.0000361906718468 ***
## Vintage2010                          0.0000023307261142 ***
## Vintage2011                          0.0000023701087761 ***
## Vintage2012                          0.0000000430810318 ***
## Vintage2013                          0.0000003041891307 ***
## Vintage2014                          0.0000000069285996 ***
## Vintage2015                          0.0000015971637571 ***
## Vintage2016                          0.0000000000448935 ***
## Vintage2017                          0.0000015186734923 ***
## Preqin_mod$ImpactDUM                           0.05935 .
## Preqin_mod$Early_StageDUM                      0.65290
## Preqin_mod$GrowthDUM                           0.62856
## Preqin_mod$BuyoutDUM                           0.22360
## Preqin_mod$Fund_of_FundsDUM                    0.03022 *
## Preqin_mod$Venture_GeneralDUM                  0.90100
## Preqin_mod$Early_Stage_SeedDUM                 0.29177
## Preqin_mod$Co_InvestmentDUM          0.0000001299357425 ***
## Preqin_mod$North_America_DUM                   0.38800
```

```
## Preqin_mod$Europe_DUM                                  0.55309
## Preqin_mod$Asia_DUM                                    0.84099
## Preqin_mod$Diversified_Multi_Regional_DUM              0.43268
## Preqin_mod$Americas_DUM                                0.02027 *
## Preqin_mod$Africa_DUM                                  0.29056
## Preqin_mod$Middle_East_and_Israel_DUM                  0.89762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.2 on 5428 degrees of freedom
## Multiple R-squared:  0.286,  Adjusted R-squared:  0.278
## F-statistic: 35.7 on 61 and 5428 DF,  p-value: <0.0000000000000002
```

```
ols.beta1 = matrix(NA, length(ols_no_int1$coefficients), 1)

# below code used to make the chart of the coeffients
for(i in 1:length(ols_no_int1$coefficients)){
 ols.beta1[i] = ols_no_int1[1]$coefficients[i]
 }

plot.ols.beta1 = data.frame( y = ols.beta1, x =
c(1:length(ols_no_int1$coefficients)))
ggplot(plot.ols.beta1, aes(x, y)) + geom_col() + theme(text =
element_text(size=12)) + ylab("Net IRR (%)") + xlab("") +
ggtitle("Regression Coefficients")
```




```
# creating a matrix for excel
```

Multiple Regression for Dummy variables with insignifant variables removed.

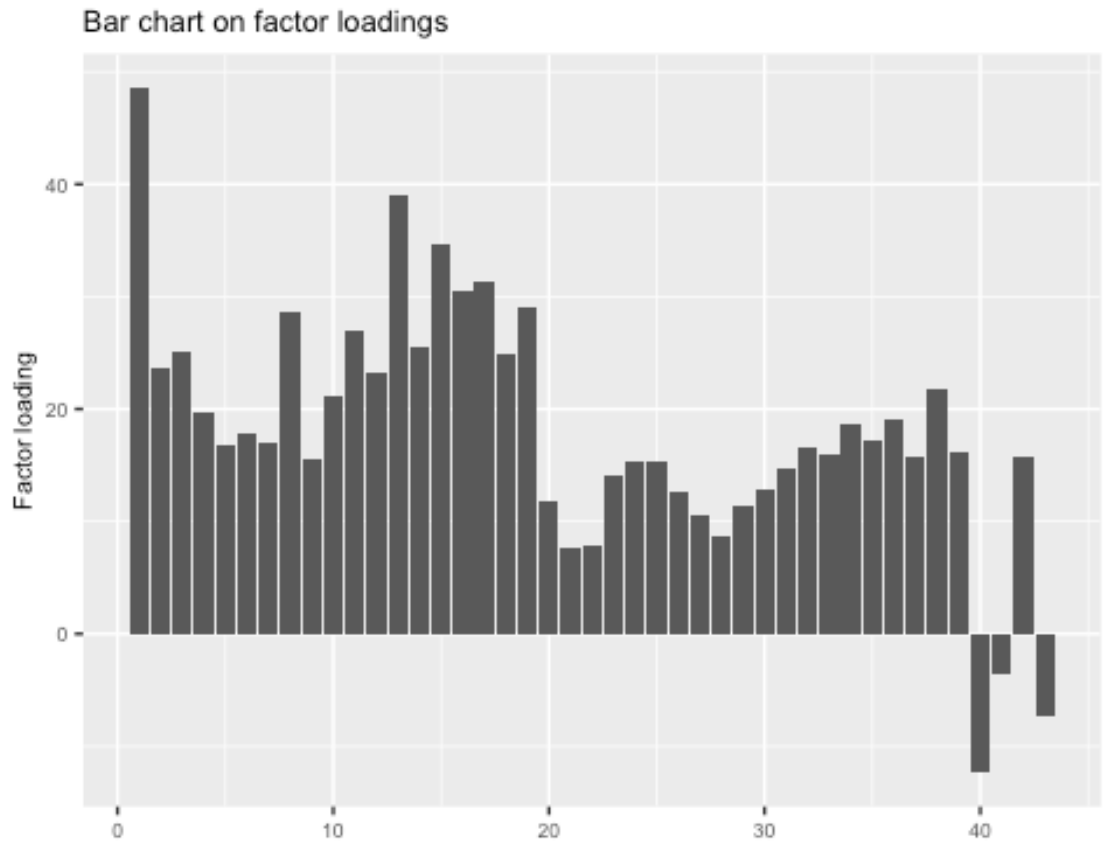
```
Vintage <- Vintage_all[,c(10,12:49)]
# above code removes insignificant Vintage Dummy variables from model 1


ols_no_int2 <- lm(Preqinclean$NET.IRR.... ~ Vintage + Preqin_mod$ImpactDUM +
Preqin_mod$Fund_of_FundsDUM + Preqin_mod$Co_InvestmentDUM +
Preqin_mod$Americas_DUM - 1)

# So far I have removed the insignificant vintage year dummy variables. Also
I have removed: Early stage, Growth, Buyout, Venture General, Early Stage
Seed strategies as well as Diversified Regionals and Africa as primary region
focus area dummy variables.

summary(ols_no_int2)

##
## Call:
## lm(formula = Preqinclean$NET.IRR.... ~ Vintage + Preqin_mod$ImpactDUM +
##     Preqin_mod$Fund_of_FundsDUM + Preqin_mod$Co_InvestmentDUM +
```

```
##     Preqin_mod$Americas_DUM - 1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -108.6   -9.4   -1.9    5.3  975.5
##
## Coefficients:
##                            Estimate Std. Error t value
Pr(>|t|)
## Vintage1978                  48.550     19.244    2.52
0.01167
## Vintage1980                  23.610      9.071    2.60
0.00928
## Vintage1981                  25.099     12.185    2.06
0.03946
## Vintage1982                  19.711      9.072    2.17
0.02985
## Vintage1983                  16.780      6.804    2.47
0.01368
## Vintage1984                  17.829      5.555    3.21
0.00134
## Vintage1985                  16.905      5.939    2.85
0.00444
## Vintage1986                  28.532      5.238    5.45
0.00000005329476938
## Vintage1987                  15.511      4.668    3.32
0.00090
## Vintage1988                  21.186      4.537    4.67
0.00000308913808779
## Vintage1989                  26.982      4.150    6.50
0.00000000008692261
## Vintage1990                  23.237      4.013    5.79
0.00000000742801260
## Vintage1991                  38.948      5.239    7.43
0.00000000000012089
## Vintage1992                  25.449      3.515    7.24
0.00000000000050774
## Vintage1993                  34.626      3.303   10.48 <
0.0000000000000002
## Vintage1994                  30.449      3.043   10.01 <
0.0000000000000002
## Vintage1995                  31.365      2.921   10.74 <
0.0000000000000002
## Vintage1996                  24.977      2.886    8.65 <
0.0000000000000002
## Vintage1997                  29.057      2.278   12.75 <
0.0000000000000002
## Vintage1998                  11.889      2.004    5.93
0.00000000318616543
## Vintage1999                   7.660      2.017    3.80
```

```
0.00015
## Vintage2000                    7.889      1.711    4.61
0.00004411966716234
## Vintage2001                   14.001      2.052    6.82
0.00000000000987810
## Vintage2002                   15.279      2.384    6.41
0.00000000015919632
## Vintage2003                   15.315      2.443    6.27
0.00000000038817913
## Vintage2004                   12.671      2.087    6.07
0.00000000136601078
## Vintage2005                   10.601      1.735    6.11
0.00000000105257089
## Vintage2006                    8.605      1.522    5.65
0.00000001659747374
## Vintage2007                   11.479      1.509    7.61
0.00000000000003300
## Vintage2008                   12.800      1.566    8.17
0.00000000000000037
## Vintage2009                   14.787      2.292    6.45
0.00000000012142755
## Vintage2010                   16.610      2.069    8.03
0.00000000000000119
## Vintage2011                   15.883      1.763    9.01 <
0.0000000000000002
## Vintage2012                   18.680      1.780   10.49 <
0.0000000000000002
## Vintage2013                   17.178      1.723    9.97 <
0.0000000000000002
## Vintage2014                   19.173      1.585   12.09 <
0.0000000000000002
## Vintage2015                   15.748      1.617    9.74 <
0.0000000000000002
## Vintage2016                   21.804      1.544   14.12 <
0.0000000000000002
## Vintage2017                   16.062      1.761    9.12 <
0.0000000000000002
## Preqin_mod$ImpactDUM         -12.308      5.976   -2.06
0.03948
## Preqin_mod$Fund_of_FundsDUM   -3.596      0.974   -3.69
0.00023
## Preqin_mod$Co_InvestmentDUM   15.662      3.001    5.22
0.00000018607751614
## Preqin_mod$Americas_DUM       -7.335      2.925   -2.51
0.01218
##
## Vintage1978                 *
## Vintage1980                 **
## Vintage1981                 *
## Vintage1982                 *
```

```
## Vintage1983                  *
## Vintage1984                  **
## Vintage1985                  **
## Vintage1986                  ***
## Vintage1987                  ***
## Vintage1988                  ***
## Vintage1989                  ***
## Vintage1990                  ***
## Vintage1991                  ***
## Vintage1992                  ***
## Vintage1993                  ***
## Vintage1994                  ***
## Vintage1995                  ***
## Vintage1996                  ***
## Vintage1997                  ***
## Vintage1998                  ***
## Vintage1999                  ***
## Vintage2000                  ***
## Vintage2001                  ***
## Vintage2002                  ***
## Vintage2003                  ***
## Vintage2004                  ***
## Vintage2005                  ***
## Vintage2006                  ***
## Vintage2007                  ***
## Vintage2008                  ***
## Vintage2009                  ***
## Vintage2010                  ***
## Vintage2011                  ***
## Vintage2012                  ***
## Vintage2013                  ***
## Vintage2014                  ***
## Vintage2015                  ***
## Vintage2016                  ***
## Vintage2017                  ***
## Preqin_mod$ImpactDUM         *
## Preqin_mod$Fund_of_FundsDUM ***
## Preqin_mod$Co_InvestmentDUM ***
## Preqin_mod$Americas_DUM      *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.2 on 5446 degrees of freedom
## Multiple R-squared:  0.284,  Adjusted R-squared:  0.278
## F-statistic: 50.2 on 43 and 5446 DF,  p-value: <0.0000000000000002

ols.beta2 = matrix(NA, length(ols_no_int2$coefficients), 1)

# below code used to make the chart of the coeffients
for(i in 1:length(ols_no_int2$coefficients)){
```

```
 ols.beta2[i] = ols_no_int2[1]$coefficients[i]
 }

plot.ols.beta2 = data.frame( y = ols.beta2, x =
c(1:length(ols_no_int2$coefficients)))
ggplot(plot.ols.beta2, aes(x, y)) + geom_col() + theme(text =
element_text(size=8)) + ylab("Factor loading") + xlab("") +
ggtitle("Bar chart on factor loadings")
```



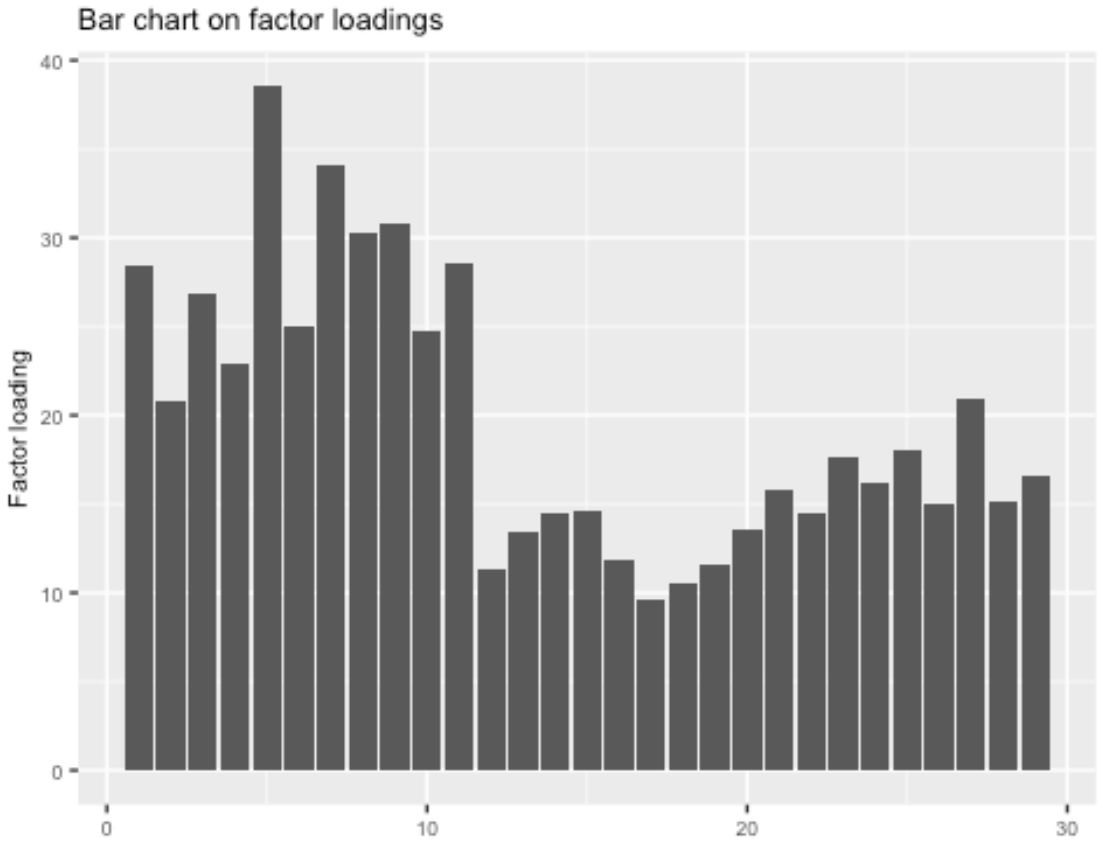
Multiple Regression for Dummy variables with all variables removed besides those of Signif. codes: 0 '***'.

```
Vintage <- Vintage_all[,c(18,20:30,33:37,39:49)]
# above code creates a subset of vintage years which had a p-value of almost
0 (*** Significance code from model 1 regression1)

ols_no_int3 <- lm(Preqinclean$NET.IRR.... ~ Vintage +
Preqin_mod$Co_InvestmentDUM - 1)

summary(ols_no_int3)

##
## Call:
```

```
## lm(formula = Preqinclean$NET.IRR.... ~ Vintage +
Preqin_mod$Co_InvestmentDUM -
##     1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -111.4   -8.4   -1.1    8.0  975.4
##
## Coefficients:
##                             Estimate Std. Error t value
Pr(>|t|)
## Vintage1986                    28.40       5.30    5.36
0.000000085639456
## Vintage1988                    20.79       4.59    4.53
0.000005966157386
## Vintage1989                    26.81       4.20    6.39
0.000000000180427
## Vintage1990                    22.92       4.06    5.65
0.000000016950283
## Vintage1991                    38.55       5.30    7.28
0.000000000000385
## Vintage1992                    25.03       3.55    7.04
0.000000000002093
## Vintage1993                    34.15       3.34   10.23 <
0.0000000000000002
## Vintage1994                    30.27       3.08    9.84 <
0.0000000000000002
## Vintage1995                    30.77       2.95   10.43 <
0.0000000000000002
## Vintage1996                    24.68       2.92    8.46 <
0.0000000000000002
## Vintage1997                    28.60       2.30   12.43 <
0.0000000000000002
## Vintage1998                    11.38       2.02    5.62
0.000000019792234
## Vintage2001                    13.37       2.07    6.46
0.000000000113084
## Vintage2002                    14.51       2.40    6.03
0.000000001714762
## Vintage2003                    14.57       2.46    5.92
0.000000003452008
## Vintage2004                    11.80       2.10    5.62
0.000000019670831
## Vintage2005                     9.62       1.74    5.54
0.000000032384118
## Vintage2007                    10.50       1.51    6.96
0.000000000003879
## Vintage2008                    11.61       1.56    7.45
0.000000000000106
## Vintage2009                    13.52       2.30    5.88
```

```
0.000000004447018
## Vintage2010                    15.82       2.08    7.60
0.000000000000035
## Vintage2011                    14.51       1.76    8.26 <
0.0000000000000002
## Vintage2012                    17.61       1.78    9.87 <
0.0000000000000002
## Vintage2013                    16.18       1.73    9.36 <
0.0000000000000002
## Vintage2014                    18.10       1.59   11.42 <
0.0000000000000002
## Vintage2015                    15.03       1.63    9.24 <
0.0000000000000002
## Vintage2016                    20.96       1.55   13.52 <
0.0000000000000002
## Vintage2017                    15.12       1.77    8.55 <
0.0000000000000002
## Preqin_mod$Co_InvestmentDUM    16.62       3.02    5.50
0.000000040323022
##
## Vintage1986                  ***
## Vintage1988                  ***
## Vintage1989                  ***
## Vintage1990                  ***
## Vintage1991                  ***
## Vintage1992                  ***
## Vintage1993                  ***
## Vintage1994                  ***
## Vintage1995                  ***
## Vintage1996                  ***
## Vintage1997                  ***
## Vintage1998                  ***
## Vintage2001                  ***
## Vintage2002                  ***
## Vintage2003                  ***
## Vintage2004                  ***
## Vintage2005                  ***
## Vintage2007                  ***
## Vintage2008                  ***
## Vintage2009                  ***
## Vintage2010                  ***
## Vintage2011                  ***
## Vintage2012                  ***
## Vintage2013                  ***
## Vintage2014                  ***
## Vintage2015                  ***
## Vintage2016                  ***
## Vintage2017                  ***
## Preqin_mod$Co_InvestmentDUM ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.5 on 5460 degrees of freedom
## Multiple R-squared:  0.266,  Adjusted R-squared:  0.262
## F-statistic: 68.2 on 29 and 5460 DF,  p-value: <0.0000000000000002

# below code used to make the chart of the coeffients
ols.beta3 = matrix(NA, length(ols_no_int3$coefficients), 1)

for(i in 1:length(ols_no_int3$coefficients)){
 ols.beta3[i] = ols_no_int3[1]$coefficients[i]
 }

plot.ols.beta3 = data.frame( y = ols.beta3, x = c(1:29))
ggplot(plot.ols.beta3, aes(x, y)) + geom_col() + theme(text =
element_text(size=8)) + ylab("Factor loading") + xlab("") +
ggtitle("Bar chart on factor loadings")
```



Checking the average of Impact and Non Impact to make sure our regression is correct.

```
Impact.idx <- which(Preqinclean$Impact...Y.N...=="Y")
meanImpactIRR <- mean(Preqinclean$NET.IRR....[Impact.idx])
print(meanImpactIRR)
```

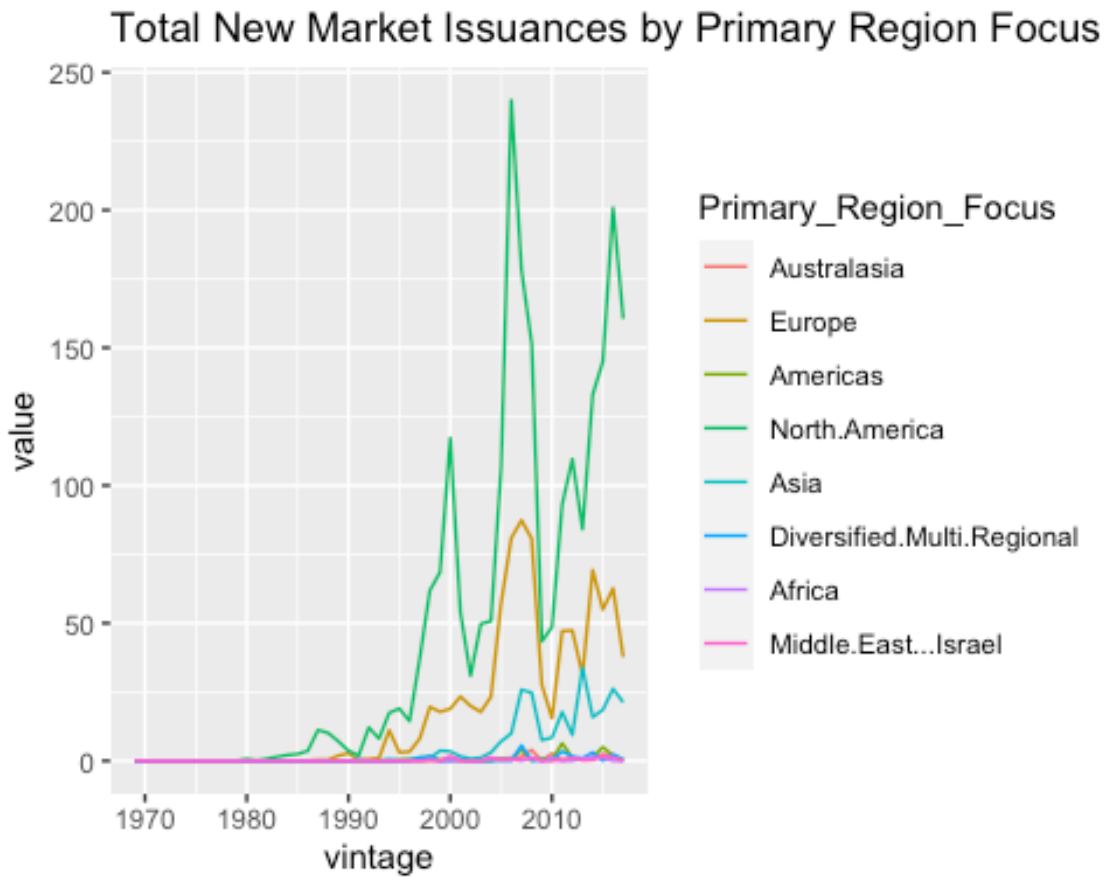
```
## [1] 3.01048

# mean of Impact IRR

nonImpact.idx <- which(Preqinclean$Impact...Y.N...=="N")
mean_nonImpactIRR <- mean(Preqinclean$NET.IRR....[nonImpact.idx])
print(mean_nonImpactIRR)

## [1] 15.6843

# mean of non Impact IRR (every fund that is not impact)

print(mean_nonImpactIRR - meanImpactIRR)

## [1] 12.6738

# difference between the two.
# In most of our regressions the impact dummy was around -12. Therefore the
impact dummy must mean how much less IRR impact funds get compared to
standard funds on average

# Now lets try the same thing with "Buyout" as a strategy to make sure we
have our bearings for that correct also
Buyout.idx <- which(Preqinclean$STRATEGY=="Buyout")
meanBuyoutIRR <- mean(Preqinclean$NET.IRR....[Buyout.idx])
print(meanBuyoutIRR)

## [1] 16.5112

# mean of Buyout IRR

nonBuyout.idx <- which(Preqinclean$STRATEGY!="Buyout")
meannonBuyoutIRR <- mean(Preqinclean$NET.IRR....[nonBuyout.idx])
print(meannonBuyoutIRR)

## [1] 15.1863

# mean of non Buyout IRR

print(meanBuyoutIRR - meannonBuyoutIRR)

## [1] 1.32488

# again so the Dummy's we are getting for the Buyout Dum are around 1.5 which
probably means that the buyout dummy's (or any strategy dummy's refer to just
the premium or reduction in IRR if your fund is a Buyout fund compared to any
other fund that is not Buyout)

# The final check I will do is for the Vintage year Dummy variables.
vintage.idx <- which(Preqinclean$VINTAGE...INCEPTION.YEAR==2017)
meanvintageIRR <- mean(Preqinclean$NET.IRR....[vintage.idx])
print(meanvintageIRR)
```

```
## [1] 15.7332

# above is the mean IRR for the vintage year of x.

# I'll check below
allvintage.idx <- which(Preqinclean$VINTAGE...INCEPTION.YEAR!=2017)
meanallvintageIRR <- mean(Preqinclean$NET.IRR....[allvintage.idx])
print(meanallvintageIRR)

## [1] 15.6313

print(meanvintageIRR - meanallvintageIRR)

## [1] 0.101903
```

Vintage Geo map code.

```
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(Preqinclean$PRIMARY.REGION.FOCUS))
vin_sum <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population

x <- unique(Preqinclean$PRIMARY.REGION.FOCUS)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
sum(Preqinclean$FUND.SIZE..USD.MN.[intersect(which(Preqinclean$VINTAGE...INCE
PTION.YEAR==(1968+i))
,which(Preqinclean$PRIMARY.REGION.FOCUS==pick[j]))])/1000
}
}
# above code creates a two variable (i,j) loop to populate the vintage year
summation matrix. The vin_sum[i,j] refers to the matrix being populated. The
loop uses various which() functions which pick out certain variables. The
first which loop for i and j = 1 picks the list of African funds (by Primary
region focus) that were stared in 1969. Once this list of creates it is run
through the second which loop that picks the corresponding fund sizes. The
loop then runs again for i and j for all vintage year dates and primary
region focus areas.

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)


# Plot of the market size by Primary region focus - not cumulative.
```

```
df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Primary_Region_Focus')
plot1 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Primary_Region_Focus)) +  ggtitle("Total New Market Issuances by Primary
Region Focus USD bln(.)") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot1
```



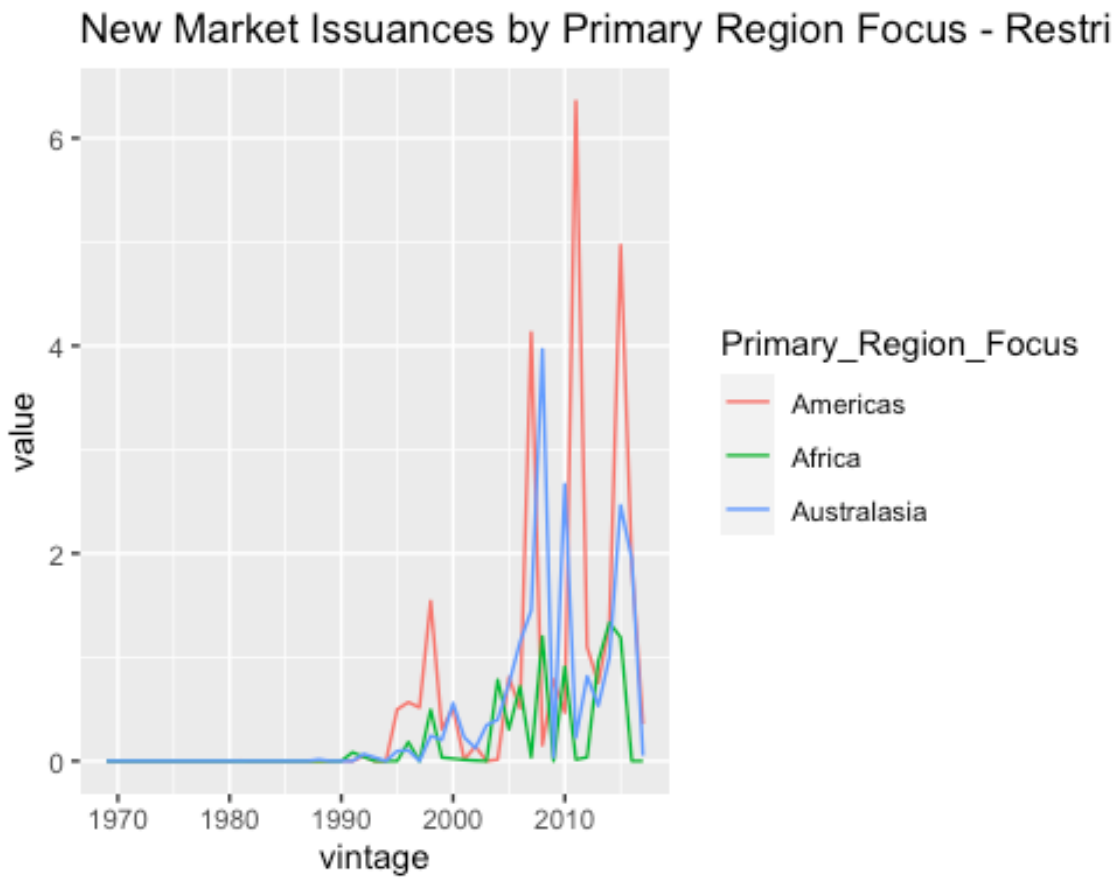
Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot1)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "1.pptx"
```

```
    )
  )
```

Vintage Geo map code restricted to smaller focus regions.

```
restricted <- c("Americas","Africa","Australasia")

min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(restricted))

vin_sum <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population

x <- unique(restricted)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
sum(Preqinclean$FUND.SIZE..USD.MN.[intersect(which(Preqinclean$VINTAGE...INCE
PTION.YEAR==(1968+i))
,which(Preqinclean$PRIMARY.REGION.FOCUS==pick[j]))])/1000
}
}

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)


# Plot of the market size by Primary region focus - not cumulative.
df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Primary_Region_Focus')
plot2 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Primary_Region_Focus)) +  ggtitle("New Market Issuances by Primary Region
Focus - Restricted - USD bln(.)") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot2
```



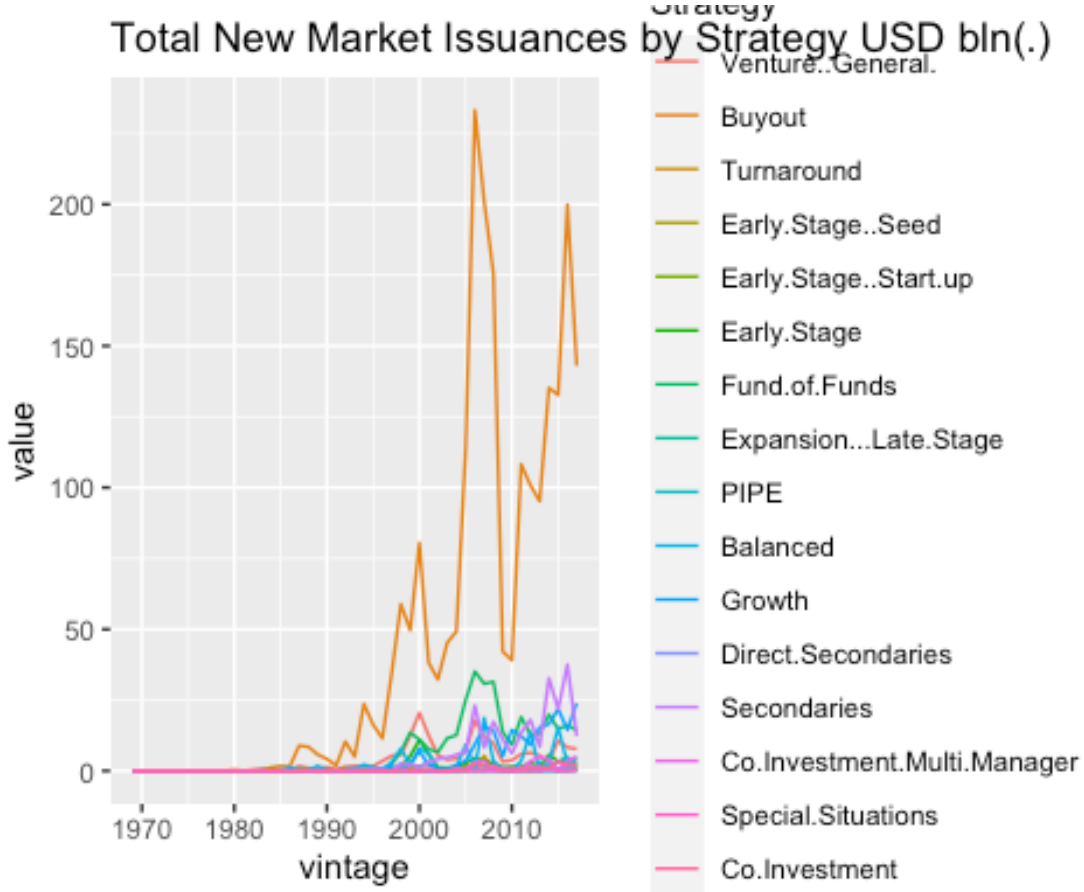
Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot2)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "2.pptx"
    )
  )
```

Vintage Strategy map code.

```
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(Preqinclean$STRATEGY))
vin_sum <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population
```

```
x <- unique(Preqinclean$STRATEGY)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
sum(Preqinclean$FUND.SIZE..USD.MN.[intersect(which(Preqinclean$VINTAGE...INCE
PTION.YEAR==(1968+i)) ,which(Preqinclean$STRATEGY==pick[j]))])/1000
}
}

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name = 'Strategy')
plot3 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour = Strategy)) +
ggtitle("Total New Market Issuances by Strategy USD bln(.)") +
theme(legend.position = "none" + scale_colour_manual(values=c(pick[1:k])))
plot3
```

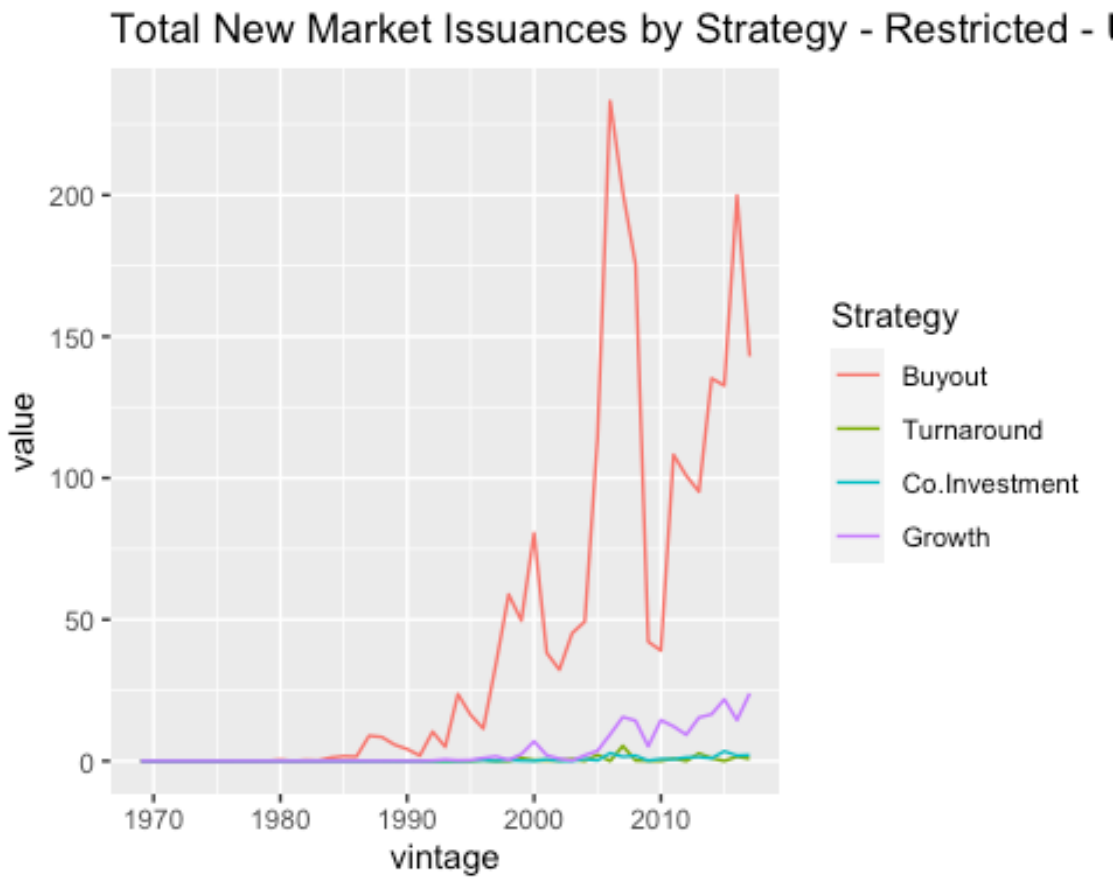


Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot3)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "3.pptx"
    )
  )
```

Vintage Strategy map code restricted.

```
restricted <- c("Buyout","Co-Investment","Growth","Turnaround")

min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(restricted))
vin_sum <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population

x <- unique(restricted)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
sum(Preqinclean$FUND.SIZE..USD.MN.[intersect(which(Preqinclean$VINTAGE...INCE
PTION.YEAR==(1968+i)) ,which(Preqinclean$STRATEGY==pick[j]))])/1000
}
}

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name = 'Strategy')
plot4 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour = Strategy)) +
ggtitle("Total New Market Issuances by Strategy - Restricted - USD bln(.)") +
theme(legend.position = "none" + scale_colour_manual(values=c(pick[1:k])))
plot4
```



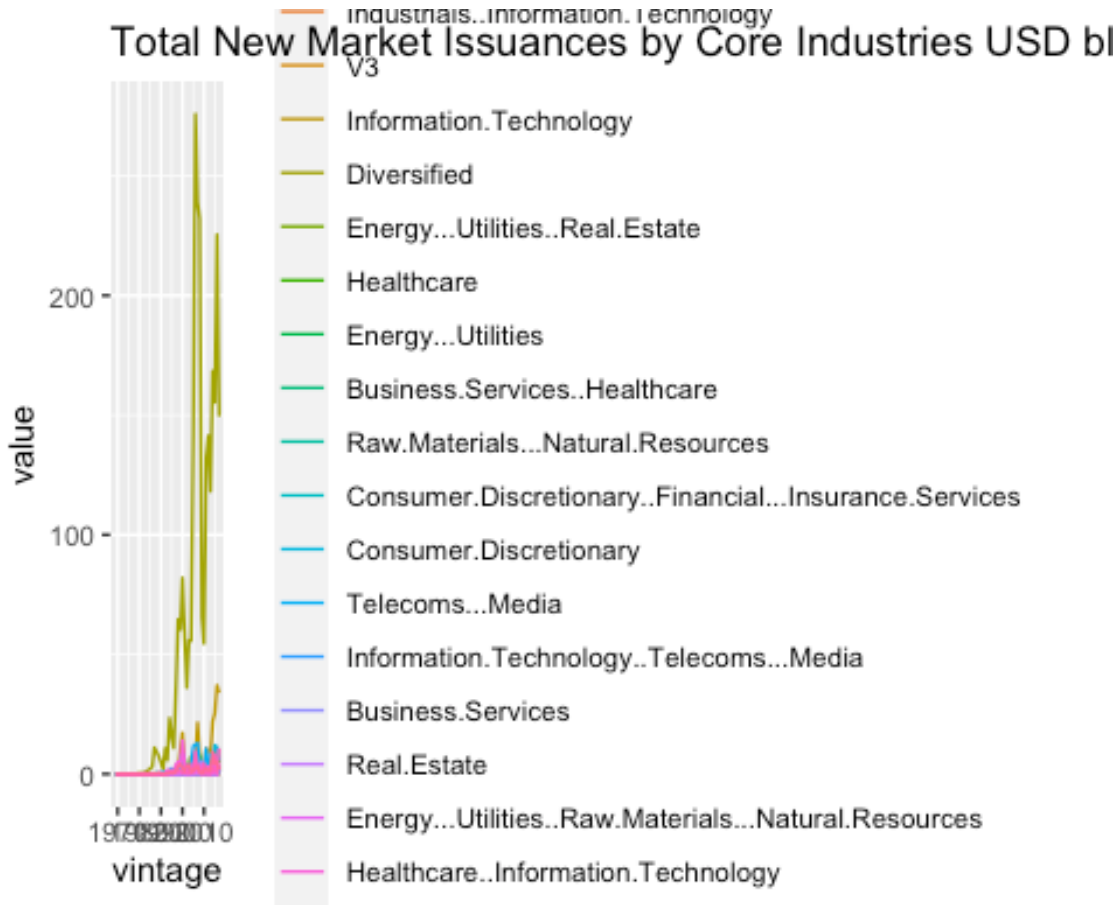
Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot4)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "4.pptx"
    )
  )
```

Vintage Core Industries map code.

```
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(Preqinclean$CORE.INDUSTRIES))
vin_sum <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population
```

```r
x <- unique(Preqinclean$CORE.INDUSTRIES)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
sum(Preqinclean$FUND.SIZE..USD.MN.[intersect(which(Preqinclean$VINTAGE...INCE
PTION.YEAR==(1968+i)) ,which(Preqinclean$CORE.INDUSTRIES==pick[j]))])/1000
}
}

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Core_Industries')
plot5 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Core_Industries)) +  ggtitle("Total New Market Issuances by Core Industries
USD bln(.)") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot5
```



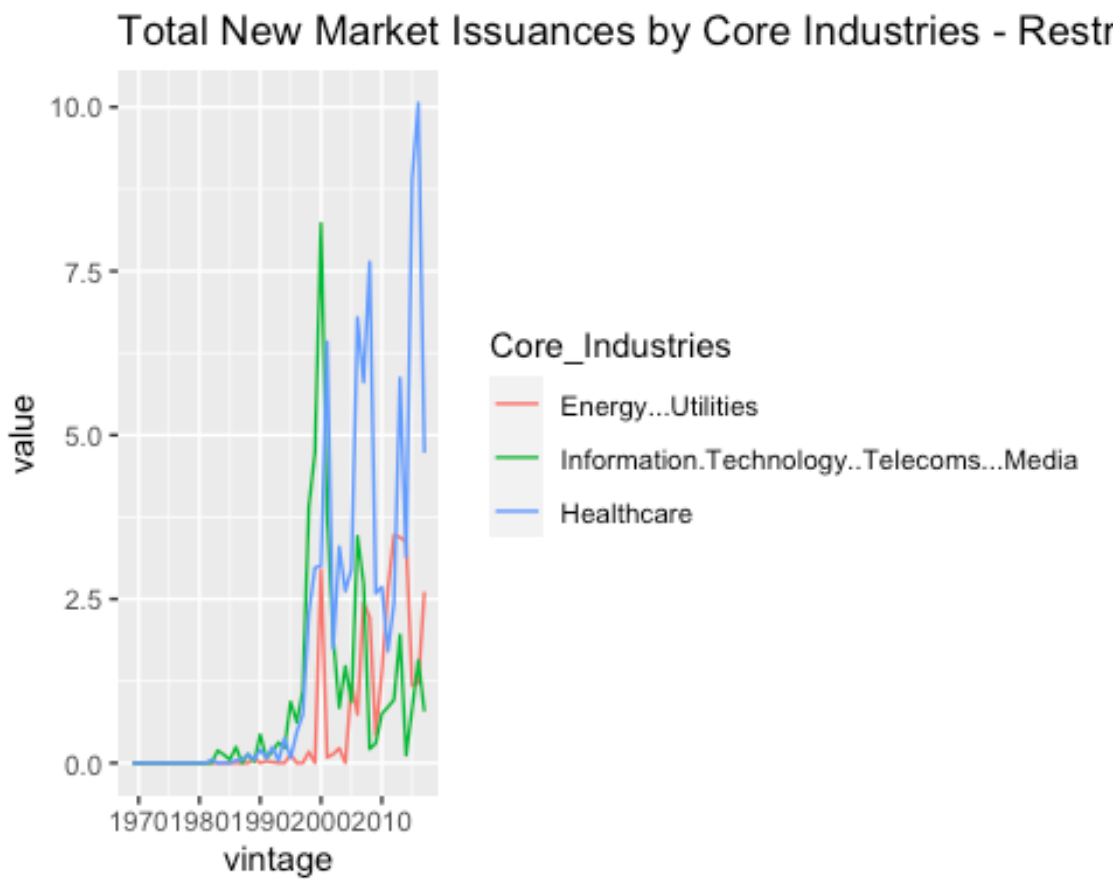
Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot5)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "5.pptx"
    )
  )
```

Vintage Core Industries map code. Restricted

```
restricted <- c("Healthcare","Energy & Utilities","Information Technology,
Telecoms & Media")

min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(restricted))
```

```r
vin_sum <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population

x <- unique(restricted)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
sum(Preqinclean$FUND.SIZE..USD.MN.[intersect(which(Preqinclean$VINTAGE...INCE
PTION.YEAR==(1968+i)) ,which(Preqinclean$CORE.INDUSTRIES==pick[j]))])/1000
}
}

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Core_Industries')
plot6 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Core_Industries)) +  ggtitle("Total New Market Issuances by Core Industries -
Restricted - USD bln(.)") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot6
```



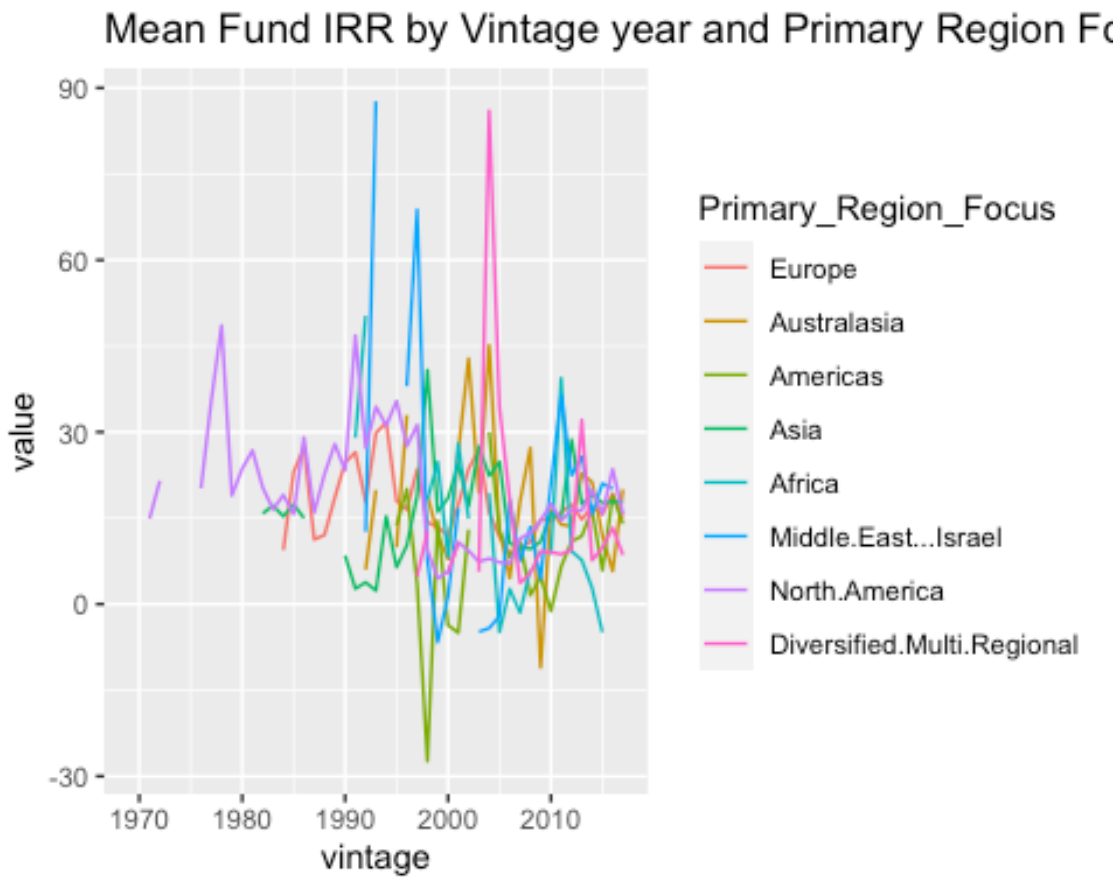
Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot6)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "6.pptx"
    )
  )
```

Vintage Geo map code by IRR. Unrestricted.

```
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(unique(Preqinclean$PRIMARY.REGION.FOCUS))
vin_mean <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population
```

```r
x <- unique(Preqinclean$PRIMARY.REGION.FOCUS)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_mean[i,j]=
mean(Preqinclean$NET.IRR....[intersect(which(Preqinclean$VINTAGE...INCEPTION.
YEAR==(1968+i)) ,which(Preqinclean$PRIMARY.REGION.FOCUS==pick[j]))])
}
}

colnames(vin_mean) <- c(pick[1:length(x)])
rownames(vin_mean) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_mean)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Primary_Region_Focus')
plot7 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Primary_Region_Focus)) +  ggtitle("Mean Fund IRR by Vintage year and Primary
Region Focus") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot7

## Warning: Removed 119 row(s) containing missing values (geom_path).
```



Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot7)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "7.pptx"
    )
  )
```

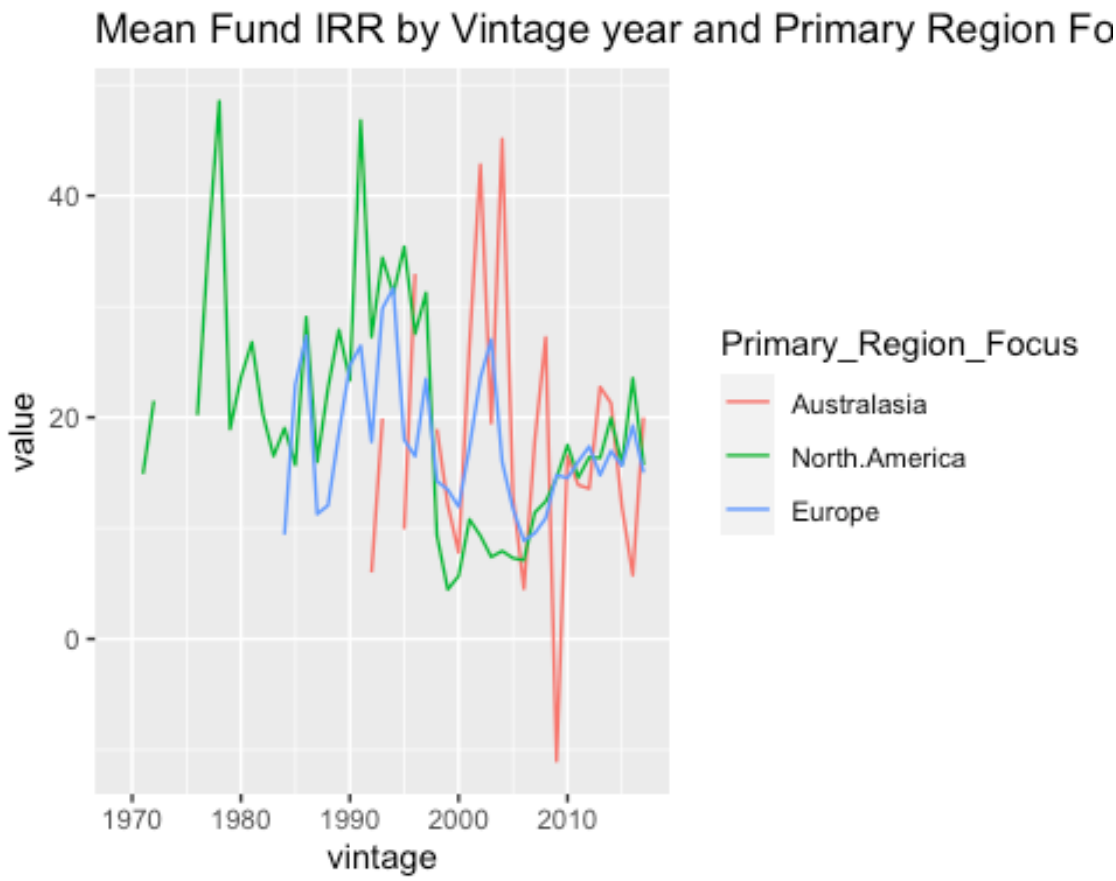
```
## Warning: Removed 119 row(s) containing missing values (geom_path).
```

Vintage Geo map code by IRR.Restricted.

```
restricted <- c("North America","Australasia","Europe")
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(restricted)
```

```r
vin_mean <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population

x <- unique(restricted)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_mean[i,j]=
mean(Preqinclean$NET.IRR....[intersect(which(Preqinclean$VINTAGE...INCEPTION.
YEAR==(1968+i)) ,which(Preqinclean$PRIMARY.REGION.FOCUS==pick[j]))])
}
}

colnames(vin_mean) <- c(pick[1:length(x)])
rownames(vin_mean) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_mean)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Primary_Region_Focus')
plot8 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Primary_Region_Focus)) +  ggtitle("Mean Fund IRR by Vintage year and Primary
Region Focus") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot8
```

```
## Warning: Removed 34 row(s) containing missing values (geom_path).
```



Export to Powerpoint

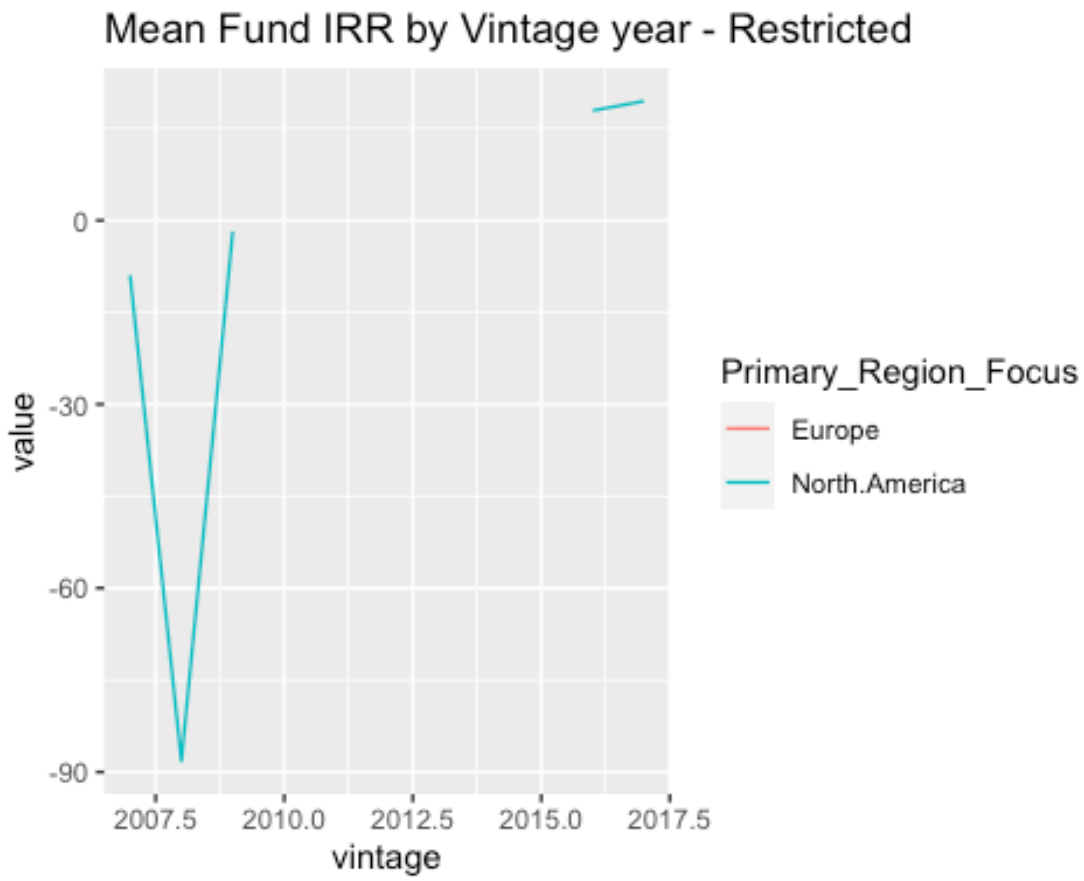
```
p_dml <- rvg::dml(ggobj = plot8)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "8.pptx"
    )
  )
```

```
## Warning: Removed 34 row(s) containing missing values (geom_path).
```

Vintage Geo map code by IRR. Restricted. Impact only.

```
restricted <- c("North America","Europe")
impact_only<- read.csv('impactonly.csv', row.names = NULL, stringsAsFactors =
FALSE)
min <- min(impact_only$VINTAGE...INCEPTION.YEAR)
```

```r
n <- max(impact_only$VINTAGE...INCEPTION.YEAR)-min + 1
k <- length(restricted)
vin_mean <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population

x <- unique(restricted)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_mean[i,j]=
mean(impact_only$NET.IRR....[intersect(which(impact_only$VINTAGE...INCEPTION.
YEAR==(min-1+i)) ,which(impact_only$PRIMARY.REGION.FOCUS==pick[j]))])
}
}

colnames(vin_mean) <- c(pick[1:length(x)])
rownames(vin_mean) <- c(min:max(impact_only$VINTAGE...INCEPTION.YEAR))

df <- data.frame(vintage = min:max(impact_only$VINTAGE...INCEPTION.YEAR),
vin_mean)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name =
'Primary_Region_Focus')
plot9 <-ggplot(df, aes(vintage,value)) + geom_line(aes(colour =
Primary_Region_Focus)) +  ggtitle("Mean Fund IRR by Vintage year -
Restricted") + theme(legend.position = "none" +
scale_colour_manual(values=c(pick[1:k])))
plot9

## Warning: Removed 4 row(s) containing missing values (geom_path).
```



Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot9)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "9.pptx"
    )
  )
```

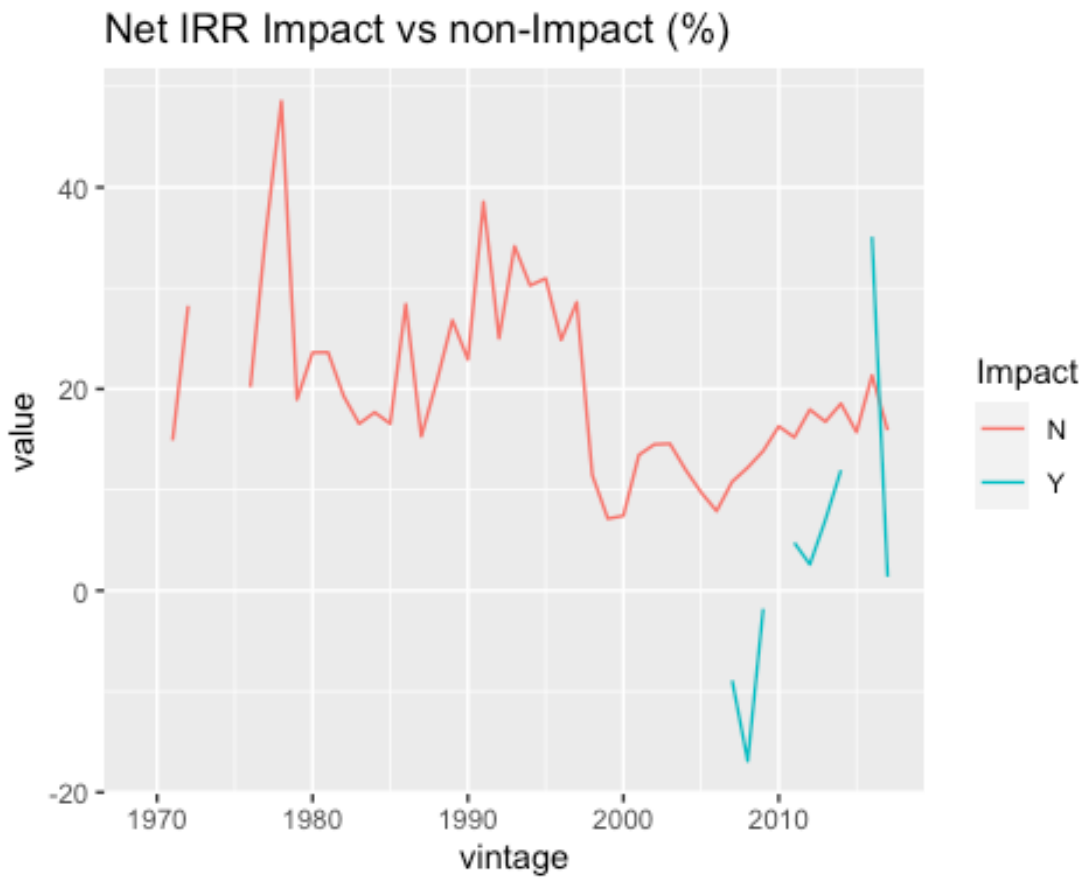
```
## Warning: Removed 4 row(s) containing missing values (geom_path).
```

Impact vs non Impact.

```
min <- min(Preqinclean$VINTAGE...INCEPTION.YEAR)
n <- max(Preqinclean$VINTAGE...INCEPTION.YEAR)-min + 1
k <- 2
vin_sum <- matrix(NA, nrow=n, ncol=k)
```

```r
# base matrix for later population

x <- c("N","Y")
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.

for(j in 1:k){
for(i in 1:n)
  {
    vin_sum[i,j]=
mean(Preqinclean$NET.IRR....[intersect(which(Preqinclean$VINTAGE...INCEPTION.
YEAR==(1968+i)) ,which(Preqinclean$Impact...Y.N...==pick[j]))])
}
}

colnames(vin_sum) <- c(pick[1:length(x)])
rownames(vin_sum) <- c(1969:2017)

df <- data.frame(vintage = 1969:2017, vin_sum)
df <- reshape2::melt(df, id.vars = 'vintage',variable.name = 'Impact')
plot10 <- ggplot(df, aes(vintage,value)) + geom_line(aes(colour = Impact)) +
ggtitle("Net IRR Impact vs non-Impact (%)") + theme(legend.position = "none"
+ scale_colour_manual(values=c(pick[1:k])))
plot10
```

```
## Warning: Removed 38 row(s) containing missing values (geom_path).
```



Export to Powerpoint

```
p_dml <- rvg::dml(ggobj = plot10)
# initialize PowerPoint slide ----
officer::read_pptx() %>%
  # add slide ----
  officer::add_slide() %>%
  # specify object and location of object ----
  officer::ph_with(p_dml, ph_location()) %>%
  # export slide -----
  base::print(
    target = here::here(
      "10.pptx"
    )
  )
```

```
## Warning: Removed 38 row(s) containing missing values (geom_path).
```

Estimation.

```
est_tableIN<- read.csv('est_tableIN.csv', row.names = NULL, stringsAsFactors
= FALSE)

colnames(est_tableIN) <- c("Estimate No.", "Fund Size (US mn)","Vintage
```

```
Year","Impact Fund?","Strategy","Prim Focus","Estimate")

coeff_table <- matrix(NA, nrow=length(names(ols_no_int1$coefficients)),
ncol=4)
# matrix for coefficent names

colnames(coeff_table) <- c("Coefficient Name","Coefficient Model 1
Value","Coefficient Pr(>|t|)","Signif. code: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
‘.’ 0.1 ‘ ’ 1")
rownames(coeff_table) <- 1:length(names(ols_no_int1$coefficients))

coeff_table[,1] <- names(ols_no_int1$coefficients)
coeff_table[,2] = ols_no_int1$coefficients
coeff_table[,3] = summary(ols_no_int1)$coefficients[,4]

for(i in 1:length(names(ols_no_int1$coefficients)))
{
    if((coeff_table[i,3]<0.1) && (coeff_table[i,3]>0.05) )#regime1end is the
last date where the regime is 1
    {
      coeff_table[i,4]="."
    }
}

for(i in 1:length(names(ols_no_int1$coefficients)))
{
    if((coeff_table[i,3]<0.05) && (coeff_table[i,3]>0.01) )#regime1end is the
last date where the regime is 1
    {
      coeff_table[i,4]="*"
    }
}

for(i in 1:length(names(ols_no_int1$coefficients)))
{
    if((coeff_table[i,3]<0.01) && (coeff_table[i,3]>0.001) )#regime1end is
the last date where the regime is 1
    {
      coeff_table[i,4]="**"
    }
}

for(i in 1:length(names(ols_no_int1$coefficients)))
{
    if(coeff_table[i,3]<0.001)
    {
      coeff_table[i,4]="***"
    }
}
```

```r
for(i in 1:length(names(ols_no_int1$coefficients)))
{
    if(coeff_table[i,3]>0.1)
    {
      coeff_table[i,4]=""
    }
}

unique(coeff_table[,1])
```

```
##  [1] "Preqinclean$FUND.SIZE..USD.MN."
##  [2] "Vintage1969"
##  [3] "Vintage1971"
##  [4] "Vintage1972"
##  [5] "Vintage1976"
##  [6] "Vintage1977"
##  [7] "Vintage1978"
##  [8] "Vintage1979"
##  [9] "Vintage1980"
## [10] "Vintage1981"
## [11] "Vintage1982"
## [12] "Vintage1983"
## [13] "Vintage1984"
## [14] "Vintage1985"
## [15] "Vintage1986"
## [16] "Vintage1987"
## [17] "Vintage1988"
## [18] "Vintage1989"
## [19] "Vintage1990"
## [20] "Vintage1991"
## [21] "Vintage1992"
## [22] "Vintage1993"
## [23] "Vintage1994"
## [24] "Vintage1995"
## [25] "Vintage1996"
## [26] "Vintage1997"
## [27] "Vintage1998"
## [28] "Vintage1999"
## [29] "Vintage2000"
## [30] "Vintage2001"
## [31] "Vintage2002"
## [32] "Vintage2003"
## [33] "Vintage2004"
## [34] "Vintage2005"
## [35] "Vintage2006"
## [36] "Vintage2007"
## [37] "Vintage2008"
## [38] "Vintage2009"
## [39] "Vintage2010"
```

```
## [40] "Vintage2011"
## [41] "Vintage2012"
## [42] "Vintage2013"
## [43] "Vintage2014"
## [44] "Vintage2015"
## [45] "Vintage2016"
## [46] "Vintage2017"
## [47] "Preqin_mod$ImpactDUM"
## [48] "Preqin_mod$Early_StageDUM"
## [49] "Preqin_mod$GrowthDUM"
## [50] "Preqin_mod$BuyoutDUM"
## [51] "Preqin_mod$Fund_of_FundsDUM"
## [52] "Preqin_mod$Venture_GeneralDUM"
## [53] "Preqin_mod$Early_Stage_SeedDUM"
## [54] "Preqin_mod$Co_InvestmentDUM"
## [55] "Preqin_mod$North_America_DUM"
## [56] "Preqin_mod$Europe_DUM"
## [57] "Preqin_mod$Asia_DUM"
## [58] "Preqin_mod$Diversified_Multi_Regional_DUM"
## [59] "Preqin_mod$Americas_DUM"
## [60] "Preqin_mod$Africa_DUM"
## [61] "Preqin_mod$Middle_East_and_Israel_DUM"

for(i in 1:6)
{
est_tableIN[i,7] <-
as.numeric(coeff_table[which(coeff_table[,1]=="Preqinclean$FUND.SIZE..USD.MN.
"),2])*est_tableIN[i,2] +
ols_no_int1$coefficients[which(names(ols_no_int1$coefficients)==est_tableIN[i
,3])] + if(est_tableIN[i,4]=="Y"){impact <-
ols_no_int1$coefficients[which(names(ols_no_int1$coefficients)=="Preqin_mod$I
mpactDUM")] }else{impact=0} +
ols_no_int1$coefficients[which(names(ols_no_int1$coefficients)==est_tableIN[i
,5])] +
ols_no_int1$coefficients[which(names(ols_no_int1$coefficients)==est_tableIN[i
,6])]
}



write.csv(coeff_table, "coeff_table.csv")
#save matrix
write.csv(est_tableIN, "Estimation.csv")
```

Dummy variables for Core Industries.

```
n <- length(Preqinclean$NET.IRR....)
k <- length(unique(Preqinclean$CORE.INDUSTRIES))
core_DUM <- matrix(NA, nrow=n, ncol=k)
# base matrix for later population
```

```
x <- unique(Preqinclean$CORE.INDUSTRIES)
pick <- sample(x, length(x), replace = FALSE, prob = NULL)
# above code picks random categorical data from primary region focus without
replacement.


for(j in 1:k){
for(i in 1:n)
{
    if(Preqinclean$CORE.INDUSTRIES[i] == pick[j])
    {
      core_DUM[i,j]=1
    }
    else
    {
      core_DUM[i,j]=0
    }
}
}

colnames(core_DUM) <- c(pick[1:length(x)])
```

Adding numericals

```
numericals <- matrix(NA, nrow=n, ncol=5)
for(i in 1:n)
{
    if(Preqinclean$NET.MULTIPLE..X.[i]=="n/a")
    {
      numericals[i,1]=1
    }
   else
    {
      numericals[i,1]=Preqinclean$NET.MULTIPLE..X.[i]
    }

  if(Preqinclean$RVPI....[i]=="n/a")
    {
      numericals[i,2]=0
    }
   else
    {
      numericals[i,2]=Preqinclean$RVPI....[i]
    }

  if(Preqinclean$DPI....[i]=="n/a")
    {
      numericals[i,3]=0
    }
   else
```

```
    {
      numericals[i,3]=Preqinclean$DPI....[i]
    }

  if(Preqinclean$CALLED....[i]=="n/a")
    {
      numericals[i,4]=100
    }
   else
    {
      numericals[i,4]=Preqinclean$CALLED....[i]
    }

  if(Preqinclean$MEDIAN.BENCHMARK.NET.IRR....[i]=="n/a")
    {
      numericals[i,5]=10
    }
   else
    {
      numericals[i,5]=Preqinclean$MEDIAN.BENCHMARK.NET.IRR....[i]
    }
}
```

Making the model dataset

```
Preqin_mod <- cbind(Preqin_mod,Vintage_all,core_DUM)
Preqin_mod <- as.matrix(Preqin_mod)
subset <- c(1:2,9:25,27:28,32:93)
Preqin_mod <- Preqin_mod[,subset]
Preqin_mod <- cbind(Preqin_mod,numericals)
write.csv(Preqin_mod,'Preqin_mod.csv')
write.csv(ols_no_int1$coefficients,'ols_no_int1$coefficients.csv')
```