

AAEC 6950 HW2 Q3

wunderwood

2024-09-24

Set wd, libraries

```
setwd("C:\\Users\\Will Underwood\\OneDrive\\Documents\\AAEC 6960\\HW2 Q3")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ forcats 1.0.0 ✓ readr 2.1.5
## ✓ ggplot2 3.5.1 ✓ stringr 1.5.1
## ✓ lubridate 1.9.3 ✓ tibble 3.2.1
## ✓ purrr 1.0.2 ✓ tidyr 1.3.1
```

```
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag() masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidyr)
```

Data Cleaning

Import Data

```
cereals0508 = read_csv("TradeData_9_24_2024_9_17_39.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
## dat <- vroom(...)
## problems(dat)
```

```
## Rows: 14489 Columns: 47
## — Column specification —
## Delimiter: ","
## chr (19): typeCode, freqCode, reporterISO, reporterDesc, flowCode, flowDesc,...
## dbl (21): refPeriodId, refYear, refMonth, period, reporterCode, partnerCode,...
## lgl (7): isOriginalClassification, isLeaf, isQtyEstimated, isAltQtyEstimate...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
cereals0920 = read_csv("TradeData_9_24_2024_9_19_36.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
## dat <- vroom(...)
## problems(dat)
```

```
## Rows: 4156 Columns: 47
## — Column specification —————
## Delimiter: ","
## chr (19): typeCode, freqCode, reporterISO, reporterDesc, flowCode, flowDesc,...
## dbl (21): refPeriodId, refYear, refMonth, period, reporterCode, partnerCode,...
## lgl (7): isOriginalClassification, isLeaf, isQtyEstimated, isAltQtyEstimate...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
gdpPHW = read_csv("productivity.csv")
```

```
## Rows: 618 Columns: 26
## — Column specification —————
## Delimiter: ","
## chr (15): STRUCTURE, STRUCTURE_ID, STRUCTURE_NAME, ACTION, LOCATION, Country...
## dbl (2): TIME_PERIOD, OBS_VALUE
## lgl (9): Time, Observation Value, Observation Status, UNIT_MEASURE, Unit of...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
laborForce = read_csv("laborForce.csv")
```

```
## Rows: 43 Columns: 20
## — Column specification —————
## Delimiter: ","
## chr (4): Series Name, Series Code, Country Name, Country Code
## dbl (16): 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
capitalFM = read_csv("capitalFormation.csv", na = "..")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 233 Columns: 20
## — Column specification —————
## Delimiter: ","
## chr (4): Series Name, Series Code, Country Name, Country Code
## dbl (16): 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
elevUnder5 = read_csv("elevationData.csv", skip = 4)
```

```
## New names:
## Rows: 266 Columns: 69
## — Column specification —————
## Delimiter: "," chr
## (4): Country Name, Country Code, Indicator Name, Indicator Code dbl (3): 1990,
## 2000, 2015 lgl (62): 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968,
## 1969, 1970, ...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...69`
```

get full cereal data

```
cereals0520 = rbind(cereals0508, cereals0920)
```

rename, select columns for leftjoin

```
gdpPHW = gdpPHW %>% rename(period = TIME_PERIOD, reporterISO = LOCATION, gdpPerHr = OBS_VALUE)
gdpPHW = gdpPHW %>% select(c("period", "reporterISO", "gdpPerHr"))
```

join gdpPHW to cereal data

```
intrCer1 = cereals0520 %>% left_join(gdpPHW, by = c("reporterISO"="reporterISO", "period" = "period"))
intrCer3 = intrCer1 %>% left_join(gdpPHW, by = c("partnerISO"="reporterISO", "period" = "period"))
intrCer3 = intrCer3 %>% rename(PgdpPH = gdpPerHr.y)
intrCer3 = intrCer3 %>% rename(gdpPerHr = gdpPerHr.x)
```

prepare capitalFM for pivot

```
capitalFM = select(capitalFM, !c('Series Name', 'Series Code', 'Country Name'))
capitalFM = capitalFM %>% rename(reporterISO = `Country Code`)
```

Pivot

```
capitalFMlong = capitalFM %>%
  pivot_longer(
    !reporterISO,
    names_to = "period",
    values_to = "capForm"
  )
```

make year values numeric

```
capitalFMlong$period = as.double(capitalFMlong$period)
```

check for and remove duplicate rows in capitalFMlong

```
capitalFMlong %>%
  count(reporterISO, period) %>%
  filter(n>1)
```

```
## # A tibble: 624 × 3
##   reporterISO period    n
##   <chr>         <dbl> <int>
## 1 ""           2005     5
## 2 ""           2006     5
## 3 ""           2007     5
## 4 ""           2008     5
## 5 ""           2009     5
## 6 ""           2010     5
## 7 ""           2011     5
## 8 ""           2012     5
## 9 ""           2013     5
## 10 ""          2014     5
## # i 614 more rows
```

```
capitalFMlong = capitalFMlong %>%
  group_by(reporterISO, period) %>%
  summarise(capForm = first(capForm), .groups = 'drop')
```

Merge intermediate cereal 3 with capitalFMlong

```
intrCer2 = intrCer3 %>% left_join(capitalFMlong, by = c("reporterISO", "period"))
intrCer4 = intrCer2 %>% left_join(capitalFMlong, by = c("partnerISO"= "reporterISO", "period" = "period"))
intrCer4 = intrCer4 %>% rename(capForm = capForm.x)
intrCer4 = intrCer4 %>% rename(PcapForm = capForm.y)
```

prepare laborForce for pivot

```
laborForce = laborForce %>% select(!c("Series Name", "Series Code", "Country Name"))
laborForce = laborForce %>% rename("reporterISO" = "Country Code")
```

pivot

```
laborForceLong = laborForce %>%
  pivot_longer(
    !reporterISO,
    names_to = "period",
    values_to = "labForce"
  )
```

make year numeric

```
laborForceLong$period = as.numeric(laborForceLong$period)
```

merge intermediate cereal 4 with laborForce

```
intrCer5 = intrCer4 %>% left_join(laborForceLong, by = c("reporterISO", "period"))
q3data = intrCer5 %>% left_join(laborForceLong, by = c("partnerISO" = "reporterISO", "period"))
q3data = q3data %>% rename(labForce = labForce.x)
q3data = q3data %>% rename(PlabForce = labForce.y)
q3data = q3data %>% rename(exports = primaryValue)
```

create tech ratio & capital intensity ratio vars, log(those) & log(exports)

```
q3data = q3data %>% mutate(techRatio = gdpPerHr/PgdpPH,
  capInt = capForm/labForce,
  PcapInt = PcapForm/PlabForce,
  capIntRatio = capInt/PcapInt,
  LtechRatio = log(techRatio),
  LcapIntRatio = log(capIntRatio),
  Lexports = log(exports))
```

elevation cleaning

```
elevUnder5 = elevUnder5 %>% select(c("Country Code", "2015"))
```

```
elevUnder5 = elevUnder5 %>% rename("elevU5" = "2015")
```

```
q3data = q3data %>% left_join(elevUnder5, by = c("reporterISO" = "Country Code"))
q3data = q3data %>% left_join(elevUnder5, by = c("partnerISO" = "Country Code"))
```

```
q3data = q3data %>% rename("elevU5" = "elevU5.x", "PelevU5" = "elevU5.y")
```

remove na & self-trading

```
q3data = na.omit(q3data) #omitted rows: partner country is non-oecd, thus not in labor, capital, or tech data
q3data = q3data %>% filter(reporterISO != partnerISO)
```

Q3.a.

i.

summarize

```
q3dataX = filter(q3data, flowCode == "X")
summary(q3dataX)
```

```
##      typeCode      freqCode      refPeriodId      refYear
## Length:442      Length:442      Min.   :20050101      Min.   :2005
## Class :character Class :character 1st Qu.:20090101      1st Qu.:2009
## Mode  :character Mode  :character Median :20180101      Median :2018
##                                     Mean  :20153155      Mean   :2015
##                                     3rd Qu.:20200101      3rd Qu.:2020
##                                     Max.   :20200101      Max.   :2020
##      refMonth      period      reporterCode      reporterISO
## Min.   :52      Min.   :2005      Min.   :124.0      Length:442
## 1st Qu.:52      1st Qu.:2009      1st Qu.:348.0      Class :character
## Median :52      Median :2018      Median :616.0      Mode  :character
## Mean   :52      Mean   :2015      Mean   :562.9
## 3rd Qu.:52      3rd Qu.:2020      3rd Qu.:842.0
```

```

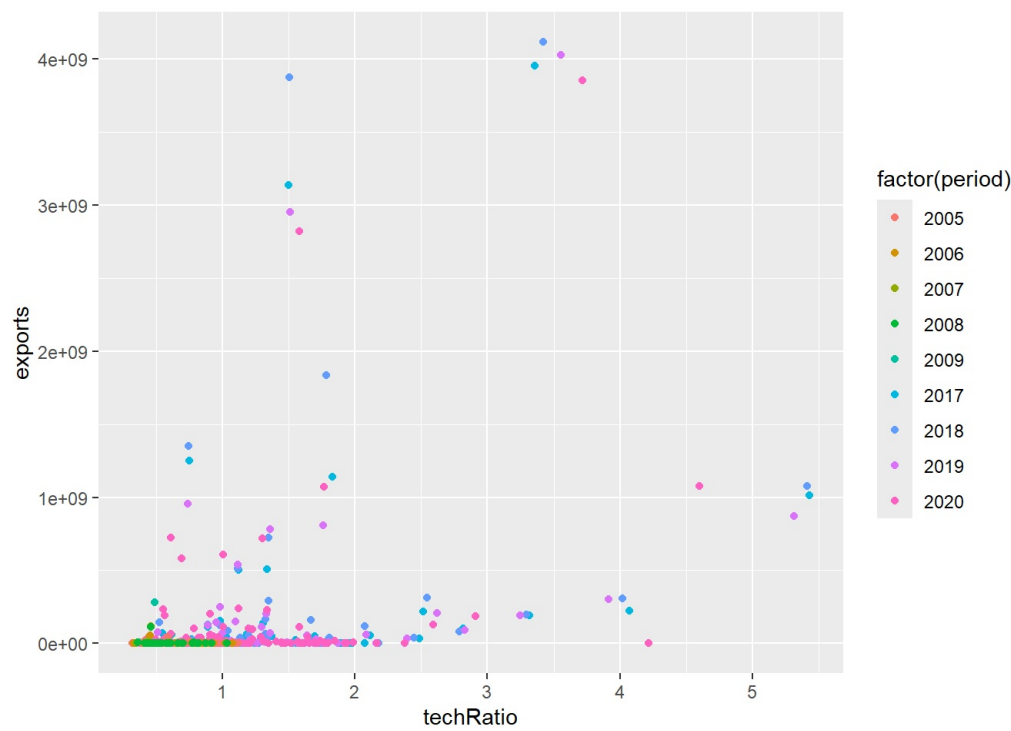
## Max. :52 Max. :2020 Max. :842.0
## reporterDesc flowCode flowDesc partnerCode
## Length:442 Length:442 Length:442 Min. : 36.0
## Class :character Class :character Class :character 1st Qu.:233.0
## Mode :character Mode :character Mode :character Median :380.0
## Mean :420.7
## 3rd Qu.:620.0
## Max. :842.0
## partnerISO partnerDesc partner2Code partner2ISO
## Length:442 Length:442 Min. :0 Length:442
## Class :character Class :character 1st Qu.:0 Class :character
## Mode :character Mode :character Median :0 Mode :character
## Mean :0
## 3rd Qu.:0
## Max. :0
## partner2Desc classificationCode classificationSearchCode
## Length:442 Length:442 Length:442
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
## isOriginalClassification cmdCode cmdDesc aggrLevel
## Mode:logical Min. :10 Length:442 Min. :2
## TRUE:442 1st Qu.:10 Class :character 1st Qu.:2
## Median :10 Mode :character Median :2
## Mean :10 Mean :2
## 3rd Qu.:10 3rd Qu.:2
## Max. :10 Max. :2
## isLeaf customsCode customsDesc mosCode motCode
## Mode :logical Length:442 Length:442 Min. :0 Min. :0
## FALSE:442 Class :character Class :character 1st Qu.:0 1st Qu.:0
## Mode :character Mode :character Median :0 Median :0
## Mean :0 Mean :0
## 3rd Qu.:0 3rd Qu.:0
## Max. :0 Max. :0
## motDesc qtyUnitCode qtyUnitAbbr qty isQtyEstimated
## Length:442 Min. :-1 Length:442 Min. :0 Mode :logical
## Class :character 1st Qu.: -1 Class :character 1st Qu.:0 FALSE:442
## Mode :character Median :-1 Mode :character Median :0
## Mean :-1 Mean :0
## 3rd Qu.: -1 3rd Qu.:0
## Max. :-1 Max. :0
## altQtyUnitCode altQtyUnitAbbr altQty isAltQtyEstimated netWgt
## Min. :-1 Length:442 Min. :0 Mode :logical Min. :0
## 1st Qu.: -1 Class :character 1st Qu.:0 FALSE:442 1st Qu.:0
## Median :-1 Mode :character Median :0 Median :0
## Mean :-1 Mean :0 Mean :0
## 3rd Qu.: -1 3rd Qu.:0 3rd Qu.:0
## Max. :-1 Max. :0 Max. :0
## isNetWgtEstimated grossWgt isGrossWgtEstimated cifvalue
## Mode :logical Min. :0 Mode :logical Min. :0
## FALSE:390 1st Qu.:0 FALSE:442 1st Qu.:0
## TRUE :52 Median :0 Median :0
## Mean :0 Mean :0
## 3rd Qu.:0 3rd Qu.:0
## Max. :0 Max. :0
## fobvalue exports legacyEstimationFlag isReported
## Min. :2.000e+00 Min. :2.000e+00 Min. :0.0000 Mode :logical
## 1st Qu.:2.823e+05 1st Qu.:2.823e+05 1st Qu.:0.0000 FALSE:442
## Median :2.382e+06 Median :2.382e+06 Median :0.0000
## Mean :1.315e+08 Mean :1.315e+08 Mean :0.4706
## 3rd Qu.:3.231e+07 3rd Qu.:3.231e+07 3rd Qu.:0.0000
## Max. :4.115e+09 Max. :4.115e+09 Max. :4.0000
## isAggregate gdpPerHr PgdpPH capForm
## Length:442 Min. :26.73 Min. : 12.86 Min. :1.011e+11
## Class :character 1st Qu.:29.32 1st Qu.: 39.89 1st Qu.:3.145e+11
## Mode :character Median :52.47 Median : 52.76 Median :1.326e+12
## Mean :50.72 Mean : 53.08 Mean :5.650e+12
## 3rd Qu.:69.88 3rd Qu.: 66.88 3rd Qu.:1.558e+13
## Max. :73.96 Max. :120.84 Max. :1.634e+13
## PcapForm labForce PlabForce techRatio
## Min. :1.198e+10 Min. : 4724407 Min. : 175622 Min. :0.3179
## 1st Qu.:1.478e+11 1st Qu.: 17594853 1st Qu.: 2760886 1st Qu.:0.6122
## Median :2.674e+11 Median : 20348244 Median : 5167188 Median :0.9426
## Mean :1.009e+12 Mean : 66521843 Mean : 17178237 Mean :1.1267
## 3rd Qu.:1.018e+12 3rd Qu.:163971527 3rd Qu.: 24586422 3rd Qu.:1.3465
## Max. :1.634e+13 Max. :167100511 Max. :167100511 Max. :5.4326
## capInt PcapInt capIntRatio LtechRatio

```

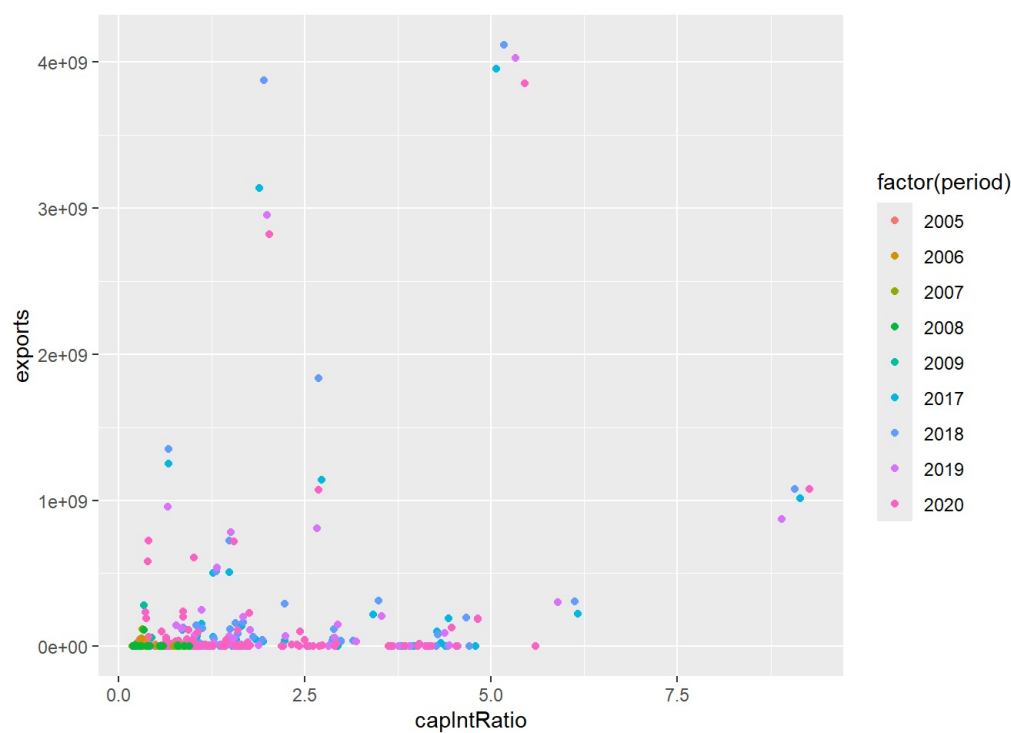
##	Min.	:15279	Min.	:10392	Min.	:0.1801	Min.	: -1.14612
##	1st Qu.	:18230	1st Qu.	:26536	1st Qu.	:0.3774	1st Qu.	: -0.49064
##	Median	:58584	Median	:57531	Median	:0.9646	Median	: -0.05907
##	Mean	:54316	Mean	:51331	Mean	:1.4463	Mean	: -0.04940
##	3rd Qu.	:95015	3rd Qu.	:67319	3rd Qu.	:1.6753	3rd Qu.	: 0.29753
##	Max.	:97806	Max.	:98736	Max.	:9.2751	Max.	: 1.69243
##	LcapIntRatio		Lexports		elevU5		PelevU5	
##	Min.	: -1.71399	Min.	: 0.6931	Min.	:0.000	Min.	: 0.000
##	1st Qu.	: -0.97440	1st Qu.	:12.5508	1st Qu.	:1.060	1st Qu.	: 0.697
##	Median	: -0.03606	Median	:14.6833	Median	:1.060	Median	: 1.599
##	Mean	: -0.06851	Mean	:14.6769	Mean	:1.229	Mean	: 3.884
##	3rd Qu.	: 0.51601	3rd Qu.	:17.2908	3rd Qu.	:1.455	3rd Qu.	: 3.282
##	Max.	: 2.22733	Max.	:22.1380	Max.	:3.312	Max.	:52.008

visualize

```
ggplot(data= q3dataX) +
  geom_point(aes(techRatio, exports, color = factor(period)))
```



```
ggplot(data = q3dataX)+
  geom_point(aes(capIntRatio, exports, color = factor(period)))
```



```
Loglog = lm(Lexports ~ LtechRatio + LcapIntRatio, data = q3dataX)
Linlog = lm(exports ~ LtechRatio + LcapIntRatio, data = q3dataX)
Linlin = lm(exports ~ techRatio + capIntRatio, data = q3dataX)
```

```
summary(Loglog)
```

```
##
## Call:
## lm(formula = Lexports ~ LtechRatio + LcapIntRatio, data = q3dataX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9913  -2.0266   0.2869   2.3732   6.8912
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.7685     0.1590  92.873 < 2e-16 ***
## LtechRatio    -0.2094     0.7763  -0.270  0.78752
## LcapIntRatio   1.4880     0.4591   3.242  0.00128 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.329 on 439 degrees of freedom
## Multiple R-squared:  0.1327, Adjusted R-squared:  0.1288
## F-statistic: 33.59 on 2 and 439 DF, p-value: 2.658e-14
```

```
summary(Linlog)
```

```
##
## Call:
## lm(formula = exports ~ LtechRatio + LcapIntRatio, data = q3dataX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -631441712 -182185679 -79623775  32516436 3608946772
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  146912566  23416348   6.274 8.45e-10 ***
## LtechRatio   487208447  114312524   4.262 2.48e-05 ***
## LcapIntRatio -125748721  67598697  -1.860  0.0635 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 490200000 on 439 degrees of freedom
## Multiple R-squared:  0.105, Adjusted R-squared:  0.1009
## F-statistic: 25.75 on 2 and 439 DF, p-value: 2.656e-11
```

```
summary(Linlin)
```

```
##
## Call:
## lm(formula = exports ~ techRatio + capIntRatio, data = q3dataX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.025e+09 -1.520e+08 -5.040e+07  3.622e+07  3.630e+09
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -199323769  46431394  -4.293 2.17e-05 ***
## techRatio    390200783  78221502   4.988 8.79e-07 ***
## capIntRatio  -75279392  40608367  -1.854  0.0644 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 477600000 on 439 degrees of freedom
## Multiple R-squared:  0.1505, Adjusted R-squared:  0.1466
## F-statistic: 38.89 on 2 and 439 DF, p-value: 2.832e-16
```

Q3.d

```
Linlin2 = lm(exports ~ techRatio + capIntRatio + elevU5 , data = q3dataX)
```

```
summary(Linlin2)
```

```
##
## Call:
## lm(formula = exports ~ techRatio + capIntRatio + elevU5, data = q3dataX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -971231774 -140098209 -52655639  33026484 3635210282
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -161593311   55909123  -2.890  0.00404 **
## techRatio    401004354   78688317   5.096 5.17e-07 ***
## capIntRatio  -76579850   40601138  -1.886  0.05994 .
## elevU5       -39077500   32293856  -1.210  0.22691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 477300000 on 438 degrees of freedom
## Multiple R-squared:  0.1533, Adjusted R-squared:  0.1475
## F-statistic: 26.44 on 3 and 438 DF, p-value: 9.809e-16
```