GROUP 01

TEAM NAME: INFERENTIA

STATE OF THE PROBLEM

The nominal GDP is the measure of country's economic output for a calendar year, using the current rates and excluding the inflation rate. For example, the real GDP of the country is 5.0 % and the inflation rate is 2.0% then the nominal GDP of the country is 3.0 %. To calculate the Nominal GDP, the data of the variables associated with it should be available. Currently, the nominal GDP data by industry is available with a 3-year-lag. While, the data associated with the variable is available at many data sources.

There are three generally accepted ways to calculate GDP:

1. Product approach: adding up the market values of all final goods/services.
2. Expenditure approach: adding up the total expenditure of different sectors of the economy.
3. Income approach: adding up the income generated by the production of final goods/services

We here, follow the expenditure based approach. The formula to calculate the GDP is given below:

GDP = C + I + G + NX

C: Consumption expenditure is the expenditure by households on consuming goods/services

I: Investment is the purchase of new capital goods (tools, instruments, machines, buildings, and other constructions) and additions to inventories

G: Government purchases is the expenditure by all levels of government on goods/services

NX: Net exports is the value of exports of goods/services minus the value of imports of goods/services

- **Exports** are goods produced within Canada and sold to the rest of the world
- **Imports** are goods produced outside Canada and purchased by Canadian households, firms, and governments

OBJECTIVE OF THIS ANALYSIS

The main objective of this analysis is to create a model which will help is predicting the Nominal GDP will a very less error rate. For creating the model, we need to use publicly available data sources, and apply data analysis techniques to generate timely estimates of nominal GDP.

DATA SETS TO BE USED FOR ANALYSIS:

[1] Multifactor productivity and related variables.
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610020801

[2] Gross domestic product (GDP) at basic prices
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610043403

[3] World Bank Data
https://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&met_y=ny_gdp_mktp_cd&idim=country:CAN:MEX:IND&hl=en&dl=en#!ctype=l&strail=false&bcs=d&nselm=h&met_y=ny_gdp_mktp_cd&scale_y=lin&ind_y=false&rdim=country&idim=country:CAN&ifdim=country&hl=en_US&dl=en&ind=false

[4] Stock Exchange Data
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1010012401

[5] Currency Exchange Rates
https://www.ofx.com/en-ca/forex-news/historical-exchange-rates/
https://www.statista.com/statistics/412804/euro-to-canadian-dollar-average-annual-exchange-rate/

Note: The other sources are all from the stats can website which will be included later.

VARIABLES

Possible Response Variables:

Variable Name	Data Years	Data Seasonality
Nominal GDP	2009 - 2016	Annual [3]

Explanatory Variables: (The ones highlighted in red have strong impact and will be used in further analysis)

Variable Name	Data Years	Data Seasonality	Indicator Type	Correlation
Population Per Year	2009-2016	Annual [3]	Leading	R = 0.18 P = 0.67
Export of goods	2010-2016	Annual [3]	Leading	R = 0.65 P = 0.81
Import of goods	2010-2016	Annual [3]	Leading	R = 0.62 P = 0.1
Real Interest Rate (%)	2009-2016	Annual [3]	Lagging	R = -0.67 P = 0.066
Domestic Companies (Total)	2009-2016	Annual [3]	Leading	R = 0.67 P = 0.088
Travel Services (% of service imports BoP)	2009-2016	Annual [3]	Leading	R = 0.9 P = 0.0022
Energy Use (Kg of oil equivalent per capita)	2009-2016	Annual [3]	Leading	R = 0.31 P = 0.45
Agricultural land (% of land area)	2009-2016	Annual [3]	Leading	R = -0.63 P = 0.093
Refugee Population	2009-2016	Annual [3]	Leading	R = 0.54

				P = 0.17
Personal Remittances (received in US $)	2009-2016	Annual [3]	Lagging	R = 0.54 P = 0.17
Personal Remittances (paid in US $)	2009-2016	Annual [3]	Leading	R = 0.94 P = 0.00049
Railways, Passengers carried	2009-2016	Annual [3]	Leading	R = -0.56 P = 0.15
Air transport, Passengers carried, Domestic	2009-2016	Annual [3]	Leading	R = 0.25 P = 0.56
Air transport, Passengers carried, Worldwide	2009-2016	Annual [3]	Leading	R = 0.1 P = 0.81
Tourists Per Year	2009-2016	Annual	Leading	R = -0.27 P = 0.53
Stock Market	2009-2016	Annual [3]	Leading	R = -0.3 P = 0.47
Toronto Stock Exchange, value of shares traded (x 1,000,000)	2009-2016	Monthly[4]	Lagging	R= -0.54 P= 0.17
Toronto Stock Exchange, volume of shares traded (x 1,000,000)	2009-2016	Monthly[4]	Lagging	R=-0.68 P=0.062
USD to CAD Exchange Rates	2009-2016	Annual	Lagging	R = -0.63 P = 0.092
10-year Government of Canada bond rate	2009-2016	Monthly[5]	Lagging	R= -0.17 p= 0.69
EUR to CAD Exchange Rates	2009-2016	Annual	Lagging	R = -0.74 P = 0.034
Housing Market	2009-2016	Annual [2]	Leading	R = 0.3 P = 0.46
Employment Rate	2009-2016	Annual	Leading	R= 0.28 P= 0.5
Ease of doing business	2009-2016	Annual	Leading	R= 0.23 P= 0.59
Incoming International Tourists	2009-2016	Annual [3]	Leading	
Real GDP	2009-2016	Annual [1]	Lagging	R = 0.28 P = 0.6
Income and Wage Growth/Decline	2009-2016	Annual [3]	Lagging	R= 0.69 p=0.059
Unemployment Rate	2009-2016	Annual [3]	Lagging	R= -0.49 P= 0.21
CPI (Inflation)	2009-2016	Annual [3]	Lagging	R = 0.56 P = 0.15
Interest Rate	2009-2016	Annual	Lagging	R= -0.68 p= 0.066
Manufacturing	2009-2016	Annual	Leading	R = 0.45 P = 0.26
Average labor productivity.	2009-2016	Annual [1]	Leading	R= 0.26 P = 0.54

Consumption	2009-2016	Annual [3]	Leading	R= 0.52 P= 0.19
Investment	2009-2016	Annual [3]	Leading	R= 0.49 P= 0.22

HYPOTHESIS AND OBSERVATION

- **We have performed EDA for almost around 35 variables and found that many show a strong positive and a strong negative correlation with the Nominal GDP of years between 2009 and 2016. (The red marked variables are the strong ones)**
- **We all add up other variables as we progress, there are few more variables identified, but as the data is monthly, we have not done the transformation and will be doing it so in the upcoming week.**
- **We have directly used values in the frames in multiple places as import files and transforming the data was time consuming and will be done in upcoming weeks.**
- **All the graphs after performing correlation are given below.**
- **After the graphs, there is non formatted rough code which has been combine from different machines.**
- **The code needs to be cleaned and automated as the values are mostly hard-coded in the frames from 2009-2016.**
- **As shown in the above table, data for export and import is available from 2010 to 2016.**
- **First we added the mean value for the year 2009, however we got much better results when we substitute it with mean.**

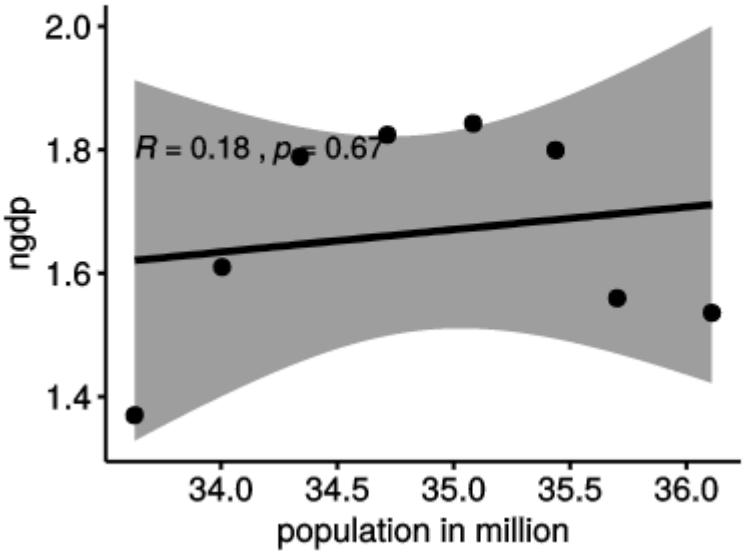
EDA ANALYSIS

```
library(plyr)
library(ggpubr)
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
ngdp <- c(1.37,1.61,1.7886,1.8243,1.8426, 1.7993,1.5596,1.5358)
population <- c(33628895.0, 34004889.0, 34339328.0, 34714222.0, 35082954.0, 35437435.0, 35702908.0, 36109487.0)
ExportAllIndustries<- c(0,359917422, 400244966, 407800730, 423614361, 469270980, 464754175,453686071)
ImportAllIndustries<- c(0,383812638, 425197604, 438058268, 451274218, 484545519, 505045626, 502117831)
RealInterestRate<- c(4.8,-0.3,-0.1,1.8,1.2,1.0,3.7,1.9)
TravelServicesImports<- c(29.1,30.5,31.1,31.3,31.1,31.0,29.9,28.7)
TotalDomesticCompanies<- c(3727,3771,3980,4030,3810,3948,3799,3368)
EnergyUse<-c(7797.1, 7788.6, 7911.6, 7733.4, 7743.7, 7897.9, 7631.3, 0)
AgrLand<-c(7,7,6.9,6.9,6.9,6.9,6.9,6.9)
RefPop<-c(99,90,109,123,99,95,80,84)
RemRec<-c(1196444724.41024,1199157750.27888, 1227344978.04621, 1254600489.98476, 1336538605.70778,
          1351304689.29191, 1274255368.12748, 1289324860.09108)
RemPaid<-c(4662098288.9915, 5290453957.41713, 5557756827.74173, 5629382995.17341, 5702499908.42515,
           5948213016.80636, 5114157603.63363, 5255722464.64191)
RailwaysPass<-c(1372.3,1346.3,1356.7,1341.8,1330.5,1300,1322.4,1380.87)
AirPass<-c(52583516,63277409.3753385,66078011.9547392,70467400.7928687,71526725.918431,75528607,80228302,85406425)
IntAirPass<-c(1198381,1234528.61925597,1245743.61579713,1280198.18557714,1263297.33686341,1290419,1322033,1359442)
dataSet<-data.frame(year,population,ExportAllIndustries, ImportAllIndustries,RealInterestRate,TotalDomesticCompanies,
                    TravelServicesImports,EnergyUse,AgrLand,RefPop,RemRec,RemPaid,RailwaysPass,AirPass,IntAirPass,ngdp)
dataSet<-dataSet%>%mutate(populationInMill=population/1000000)
dataSet<-dataSet%>%mutate(ExportMillDollar=ExportAllIndustries/1000000)
dataSet<-dataSet%>%mutate(ImportMillDollar=ImportAllIndustries/1000000)
```

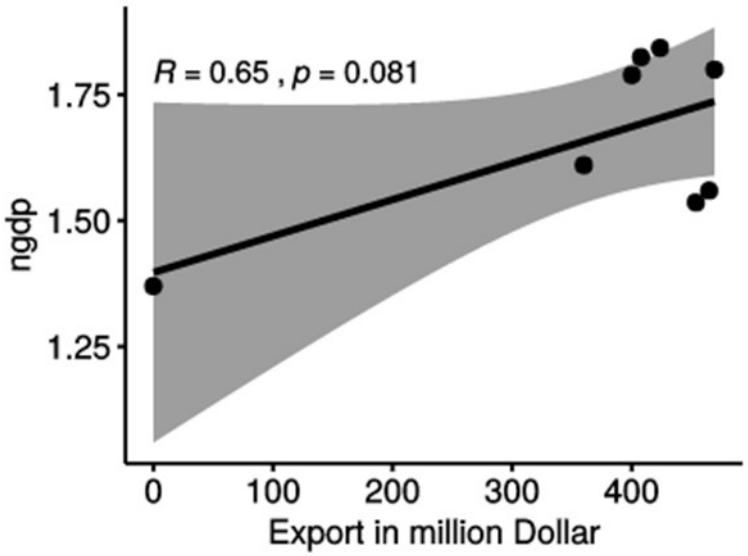
```
dataSet%>%
rename(yr=year,
       pop=population,
       exp=ExportAllIndustries,
       imp=ImportAllIndustries,
       rir=RealInterestRate,
       tsi=TravelServicesImports,
       tdc=TotalDomesticCompanies,
       engu=EnergyUse,
       agl=AgrLand,
       rp=RailwaysPass,
       ap=AirPass,
       iap=IntAirPass,
       pmd=populationInMill,
        emd=ExportMillDollar,
        imd=ImportMillDollar
       )
```

yr	pop	exp	imp	rir	tdc	tsi	engu	agl	RefPop	RemRec	RemPaid	rp	ap	iap	ngdp	pmd	emd	imd
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2009	33628895	0	0	4.8	3727	29.1	7797.1	7.0	99	1196444724	4662098289	1372.30	52583516	1198381	1.3700	33.62889	0.0000	0.0000
2010	34004889	359917422	383812638	-0.3	3771	30.5	7788.6	7.0	90	1199157750	5290453957	1346.30	63277409	1234529	1.6100	34.00489	359.9174	383.8126
2011	34339328	400244966	425197604	-0.1	3980	31.1	7911.6	6.9	109	1227344978	5557756828	1356.70	66078012	1245744	1.7886	34.33933	400.2450	425.1976
2012	34714222	407800730	438058268	1.8	4030	31.3	7733.4	6.9	123	1254600490	5629382995	1341.80	70467401	1280198	1.8243	34.71422	407.8007	438.0583
2013	35082954	423614361	451274218	1.2	3810	31.1	7743.7	6.9	99	1336538606	5702499908	1330.50	71526726	1263297	1.8426	35.08295	423.6144	451.2742
2014	35437435	469270980	484545519	1.0	3948	31.0	7897.9	6.9	95	1351304689	5948213017	1300.00	75528607	1290419	1.7993	35.43744	469.2710	484.5455
2015	35702908	464754175	505045626	3.7	3799	29.9	7631.3	6.9	80	1274255368	5114157604	1322.40	80228302	1322033	1.5596	35.70291	464.7542	505.0456
2016	36109487	453686071	502117831	1.9	3368	28.7	0.0	6.9	84	1289324860	5255722465	1380.87	85406425	1359442	1.5358	36.10949	453.6861	502.1178

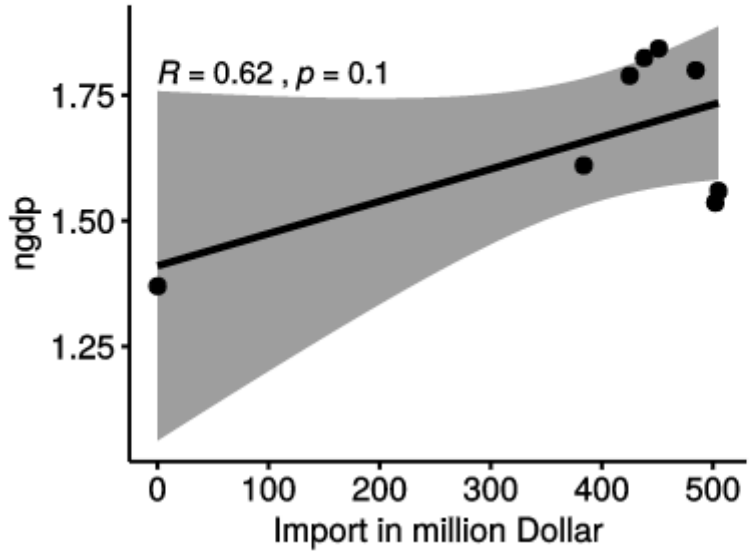
```
ggscatter(dataSet, x = "populationInMill", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "population in million", ylab = "ngdp")
```



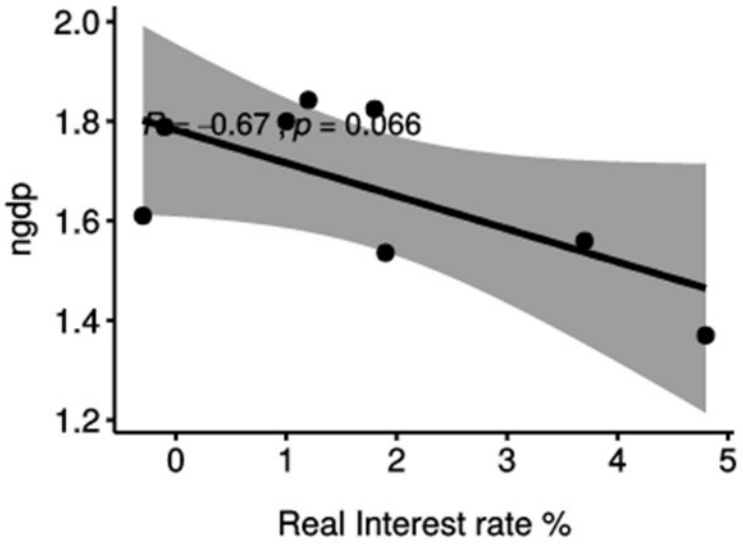

```
ggscatter(dataSet, x = "ExportMillDollar", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Export in million Dollar", ylab = "ngdp")
```



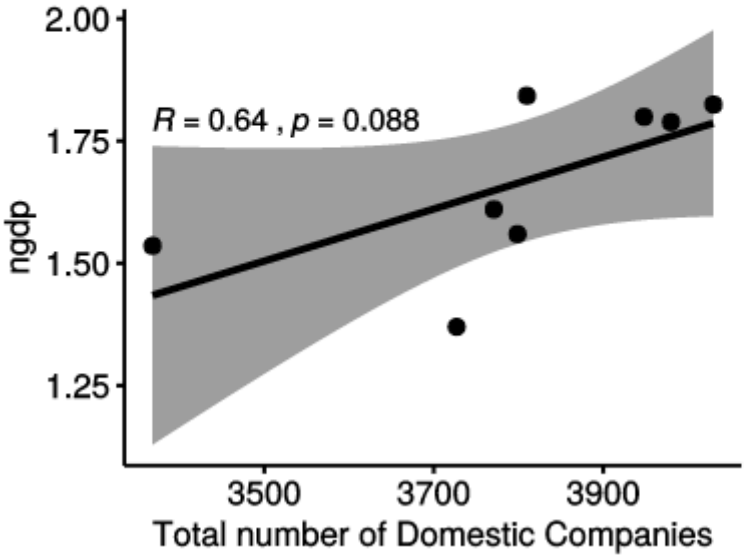
```
ggscatter(dataSet, x = "ImportMillDollar", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Import in million Dollar", ylab = "ngdp")
```



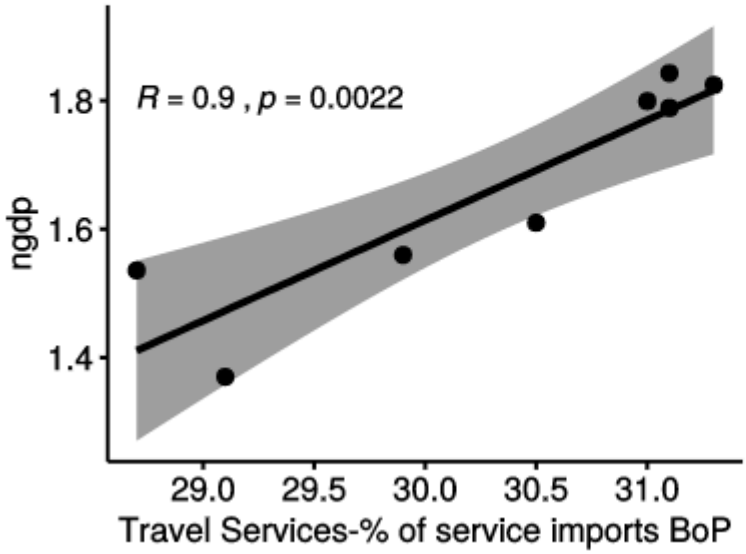

```
ggscatter(dataSet, x = "RealInterestRate", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Real Interest rate %", ylab = "ngdp")
```



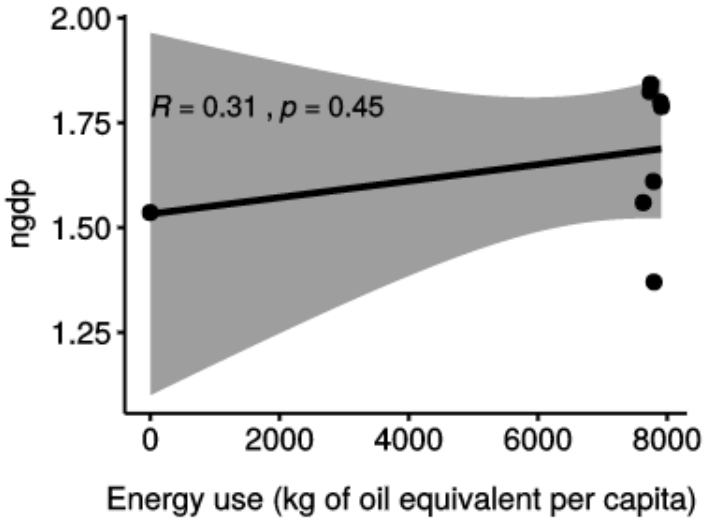
```
ggscatter(dataSet, x = "TotalDomesticCompanies", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Total number of Domestic Companies", ylab = "ngdp")
```



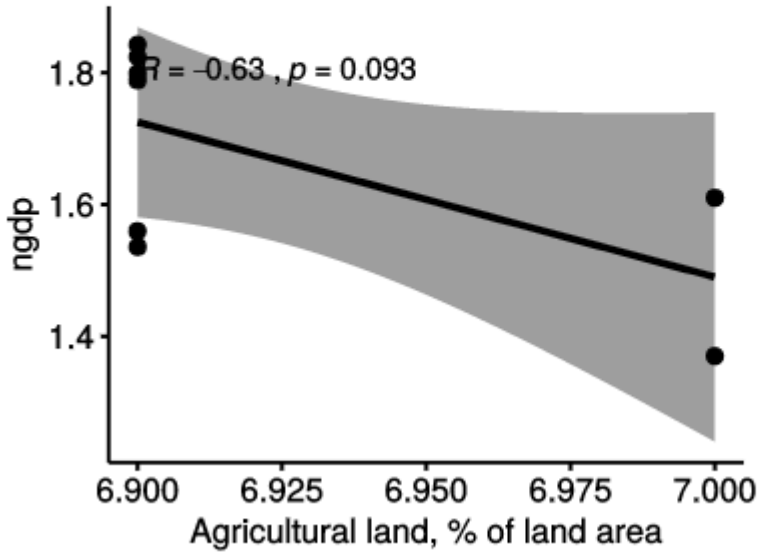

```
ggscatter(dataSet, x = "TravelServicesImports", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Travel Services-% of service imports BoP", ylab = "ngdp")
```



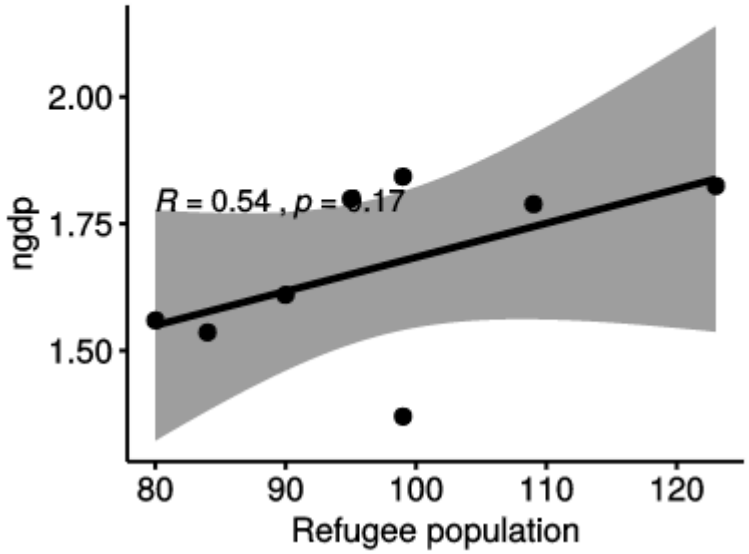
```
ggscatter(dataSet, x = "EnergyUse", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Energy use (kg of oil equivalent per capita) ", ylab = "ngdp")
```



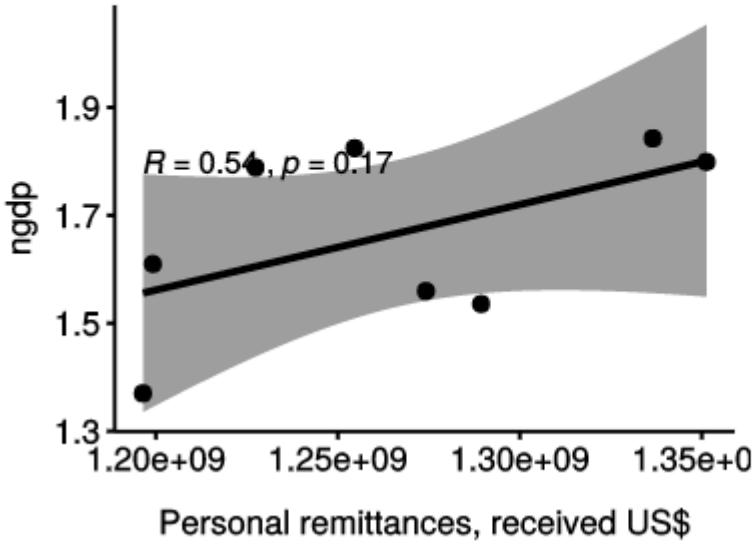

```
ggscatter(dataSet, x = "AgrLand", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Agricultural land, % of land area", ylab = "ngdp")
```



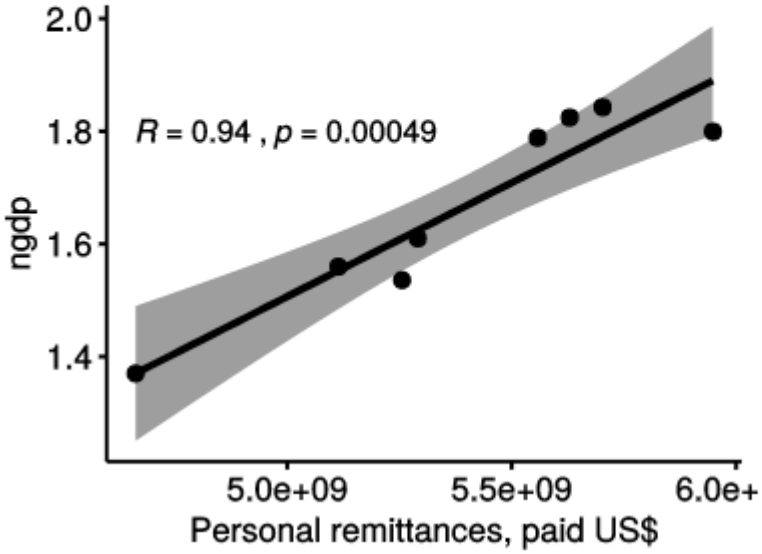
```
ggscatter(dataSet, x = "RefPop", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Refugee population", ylab = "ngdp")
```



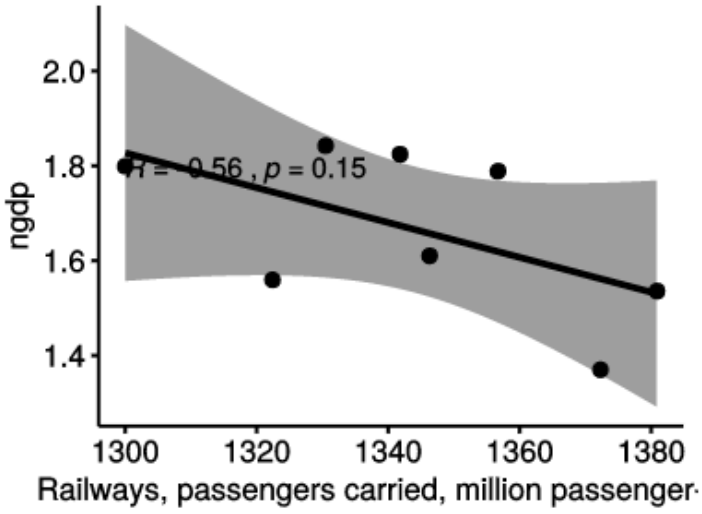

```
ggscatter(dataSet, x = "RemRec", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Personal remittances, received US$", ylab = "ngdp")
```



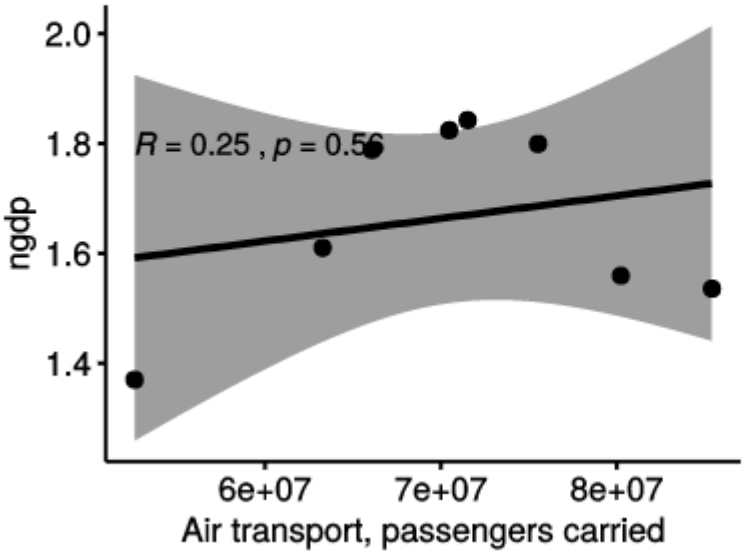
```
ggscatter(dataSet, x = "RemPaid", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Personal remittances, paid US$", ylab = "ngdp")
```



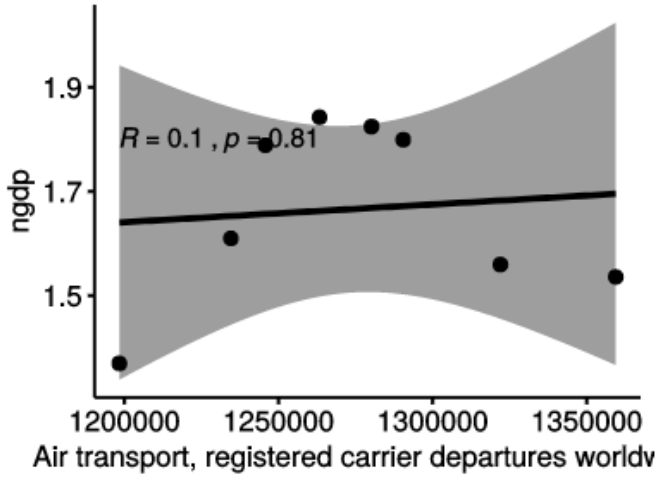

```
ggscatter(dataSet, x = "RailwaysPass", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Railways, passengers carried, million passenger-km", ylab = "ngdp")
```

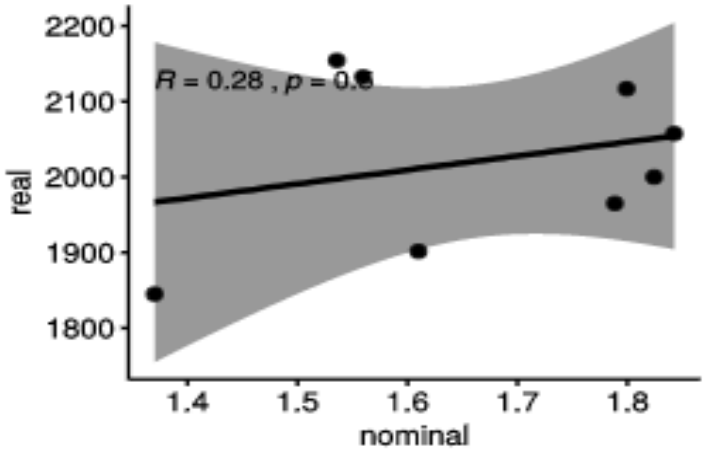


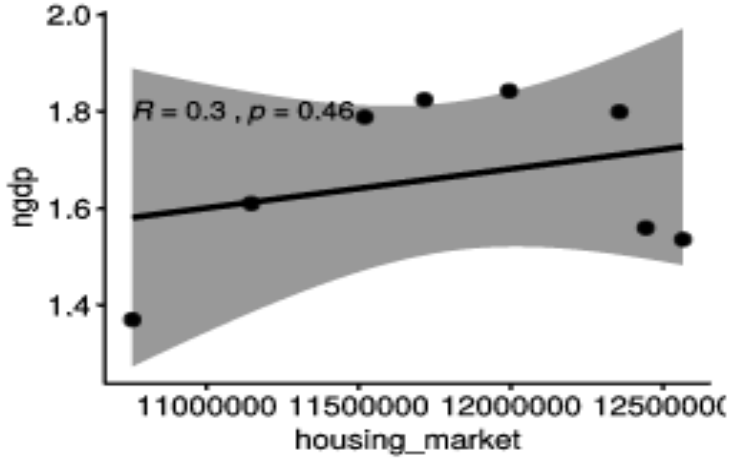
```
ggscatter(dataSet, x = "AirPass", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Air transport, passengers carried", ylab = "ngdp")
```

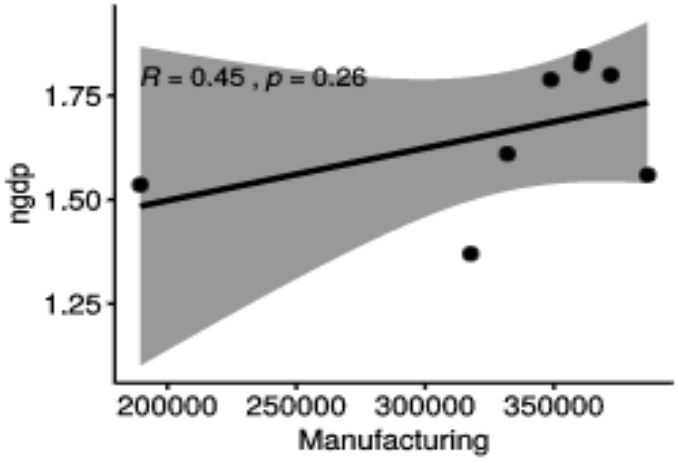


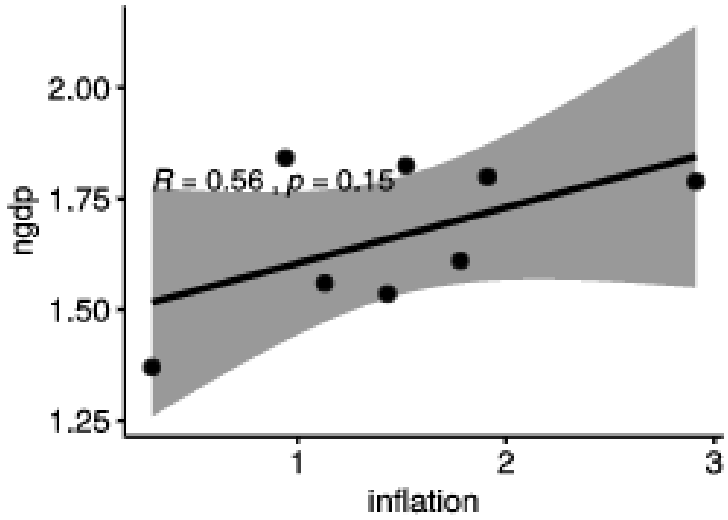

```
ggscatter(dataSet, x = "IntAirPass", y = "ngdp",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Air transport, registered carrier departures worldwide", ylab = "ngdp")
```

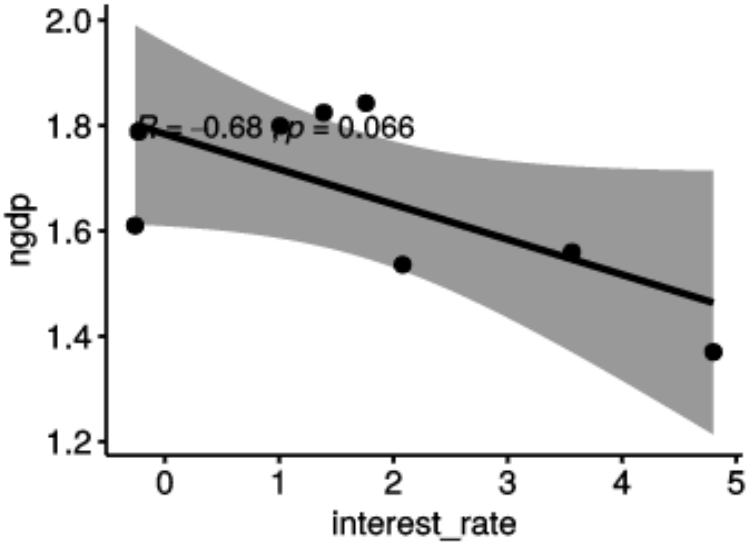



2200
2100
2000
1900
1800
real
R = 0.28 , p = 0.6
1.4
1.5
1.6
1.7
1.8
nominal


2.0
1.8
1.6
1.4
ngdp
R = 0.3 , p = 0.46
11000000
11500000
12000000
housing_market

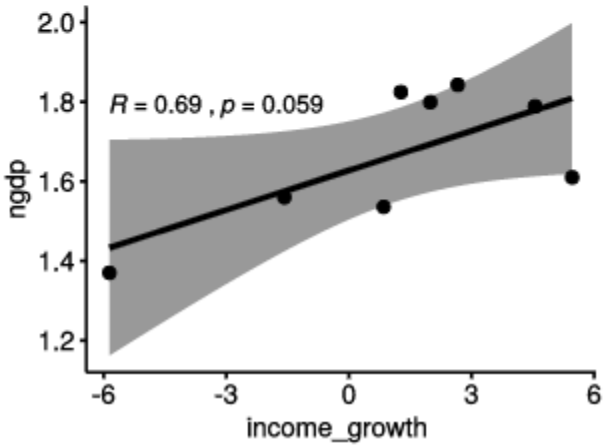

1.75
1.50
1.25
ngdp
R = 0.45 , p = 0.26
200000
250000
300000
350000
Manufacturing

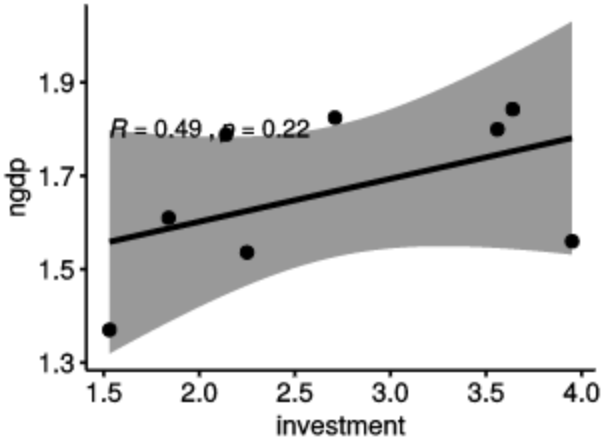

2.00
1.75
1.50
1.25
ngdp
R = 0.56 , p = 0.15
1
2
3
inflation

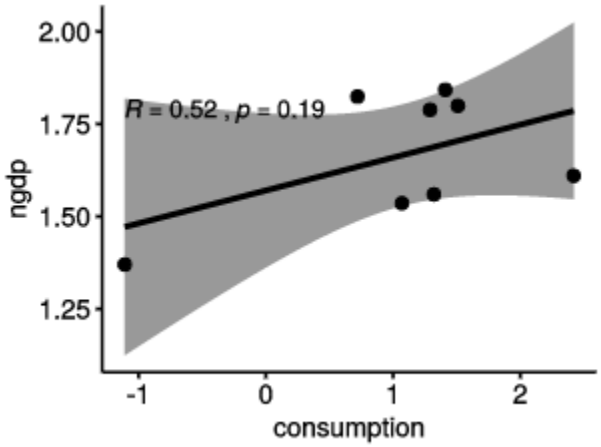

2.0
1.8
1.6
1.4
1.2
ngdp
R = −0.68 , p = 0.066
0
1
2
3
4
5
interest_rate

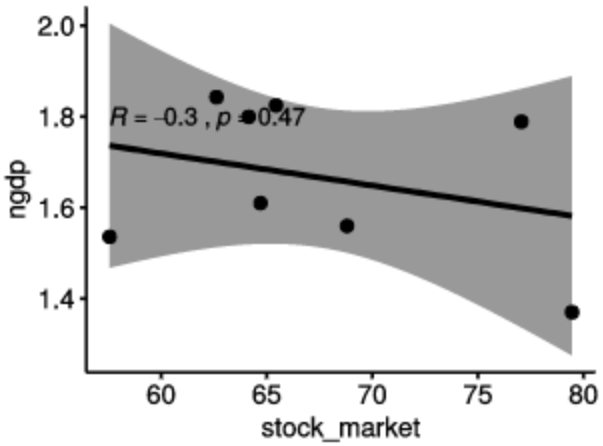
A data.frame: 8 x 8

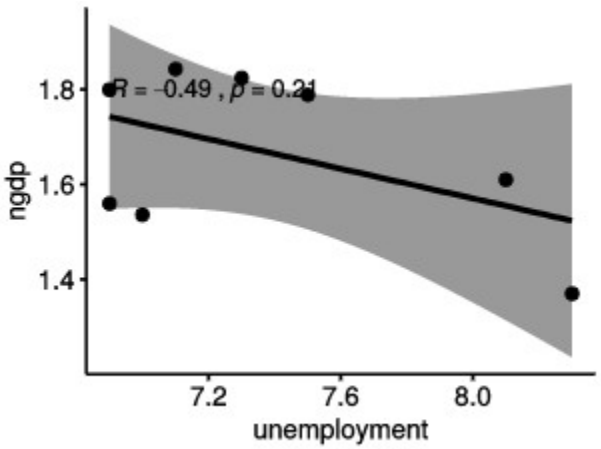
year	Manufacturing	housing_market	ngdp	labour_productivity	real_gdp	interest_rate	income_growth
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2009	317712.4	10756375	1.3700	1947.025	1845.013	4.80	-5.86
2010	331923.7	11146943	1.6100	1958.620	1901.550	-0.26	5.46
2011	348774.4	11519574	1.7886	1984.771	1965.018	-0.23	4.55
2012	360722.6	11715862	1.8243	2000.000	2000.000	1.39	1.27
2013	361340.7	11994309	1.8426	2034.268	2057.445	1.76	2.66
2014	372046.6	12356841	1.7993	2106.631	2116.766	1.01	1.99
2015	386448.9	12443428	1.5596	2109.523	2132.351	3.56	-1.58
2016	189379.7	12565782	1.5358	2130.964	2154.237	2.08	0.85

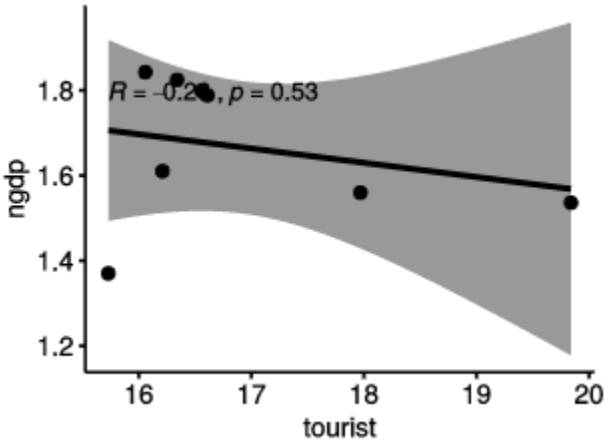


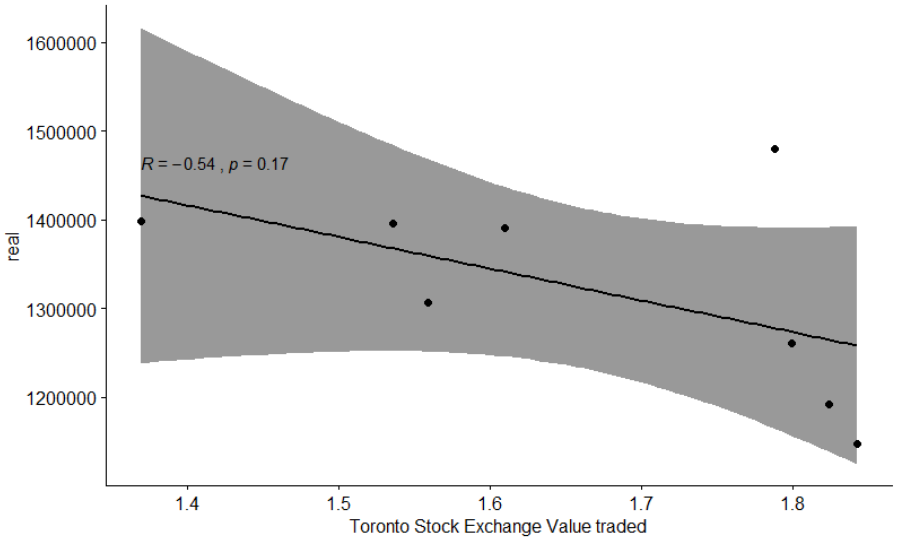


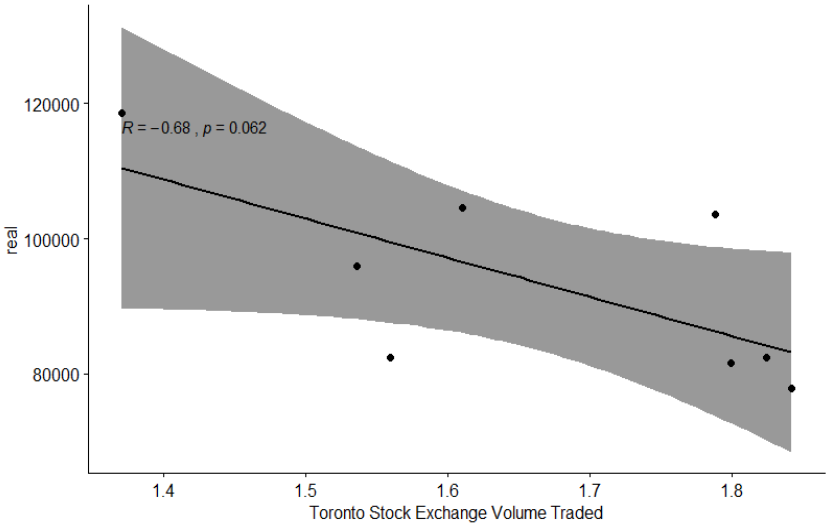

2.00
1.75
1.50
1.25
ngdp
R = 0.52 , p = 0.19
-1
0
1
2
consumption

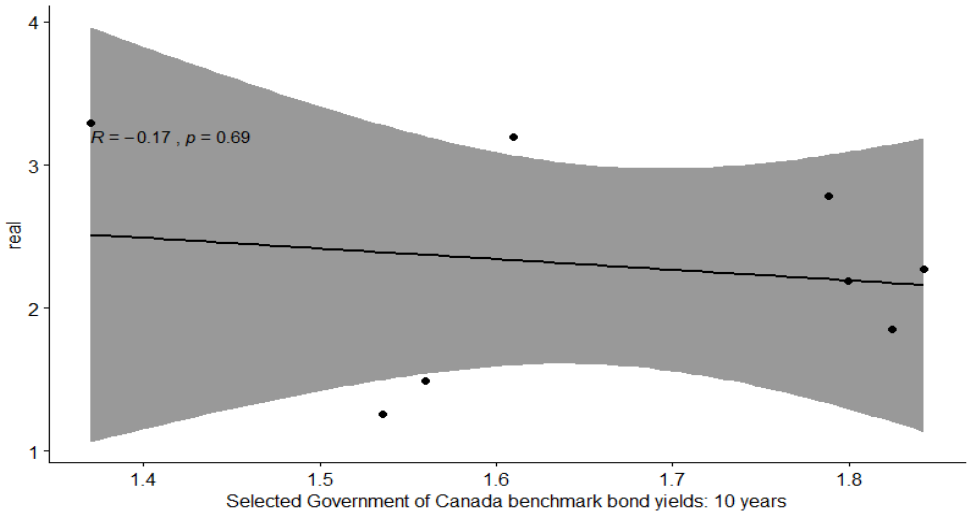

2.0
1.8
1.6
1.4
ngdp
R = −0.3 , p = 0.47
60
65
70
75
80
stock_market

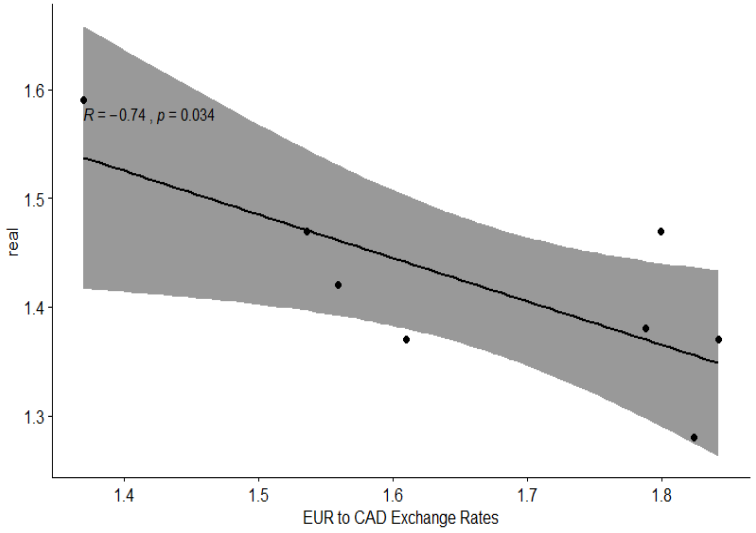

1.8
1.6
1.4
ngdp
R = −0.49 , p = 0.21
7.2
7.6
8.0
unemployment

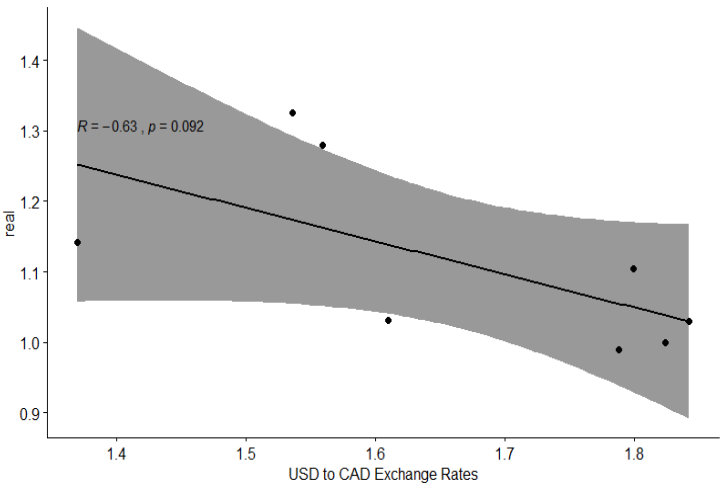

R = −0.2 , p = 0.53
ngdp
1.8
1.6
1.4
1.2
16
17
18
19
20
tourist

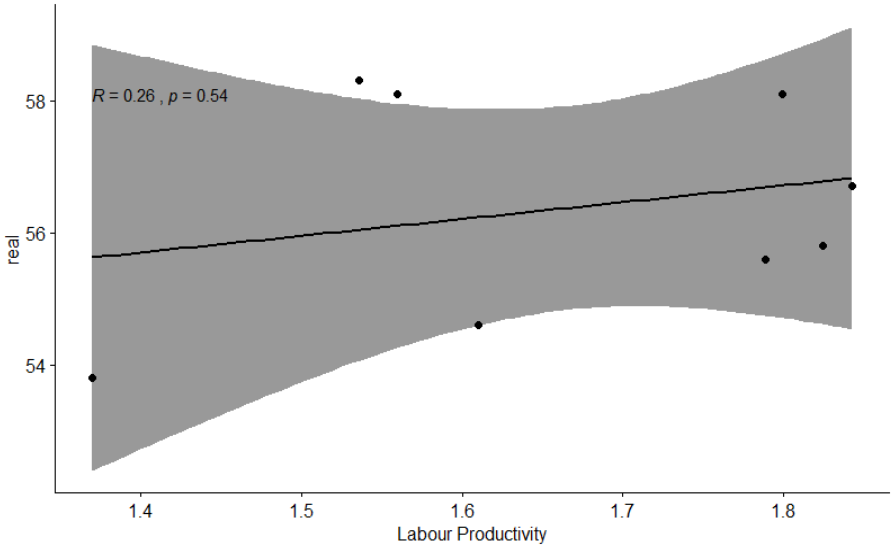

R = −0.54 , p = 0.17
real
1600000
1500000
1400000
1300000
1200000
1.4
1.5
1.6
1.7
1.8
Toronto Stock Exchange Value traded

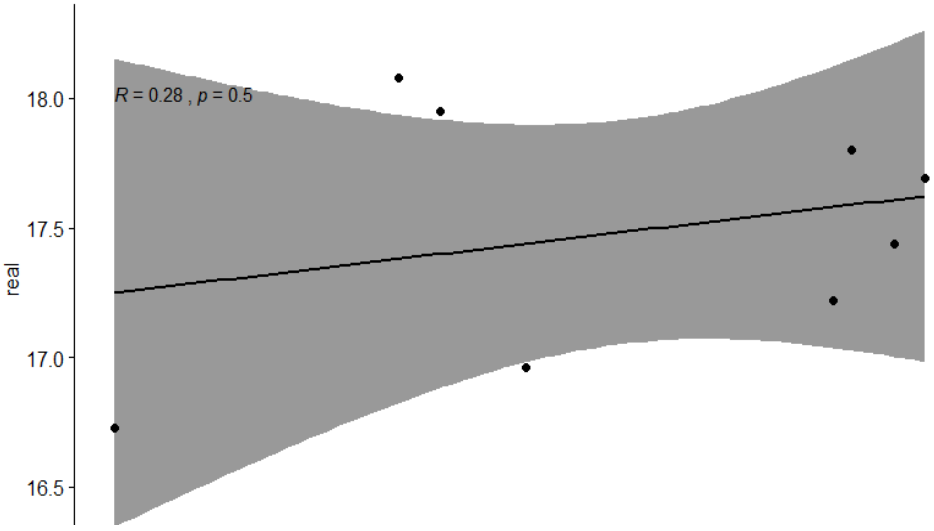

R = −0.68 , p = 0.062
real
120000
100000
80000
1.4
1.5
1.6
1.7
1.8
Toronto Stock Exchange Volume Traded

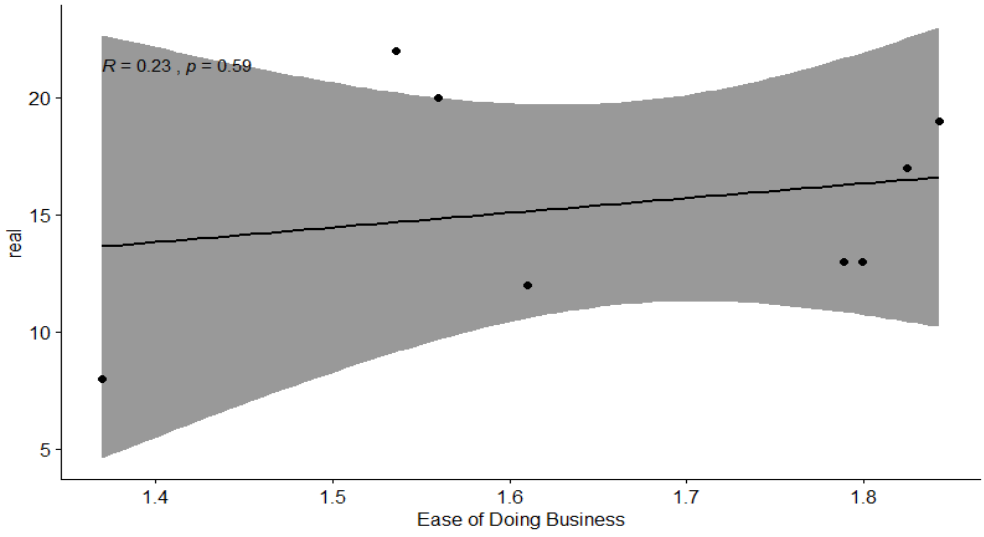

R = −0.17 , p = 0.69
real
4
3
2
1
1.4
1.5
1.6
1.7
1.8
Selected Government of Canada benchmark bond yields: 10 years


R = −0.74 , p = 0.034
real
1.6
1.5
1.4
1.3
1.4
1.5
1.6
1.7
1.8
EUR to CAD Exchange Rates


R = −0.63 , p = 0.092
real
1.4
1.3
1.2
1.1
1.0
0.9
1.4
1.5
1.6
1.7
1.8
USD to CAD Exchange Rates


R = 0.26 , p = 0.54
real
58
56
54
1.4
1.5
1.6
1.7
1.8
Labour Productivity


R = 0.28 , p = 0.5
real
18.0
17.5
17.0
16.5


R = 0.23 , p = 0.59
real
20
15
10
5
1.4
1.5
1.6
1.7
1.8
Ease of Doing Business

#nominal gdp dataframe

```
# putting nominal gdp data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
ngdp <- c(1.37,1.61,1.7886,1.8243,1.8426, 1.7993,1.5596,1.5358)
df1 <- data.frame(year, ngdp)
```

A data.frame: 8 x 2

year	ngdp
<dbl>	<dbl>
2009	1.3700
2010	1.6100
2011	1.7886
2012	1.8243
2013	1.8426
2014	1.7993
2015	1.5596
2016	1.5358

```
# preparing housing market dataframe from[2]
install.packages("readxl")
library("readxl")
my_data <- read_excel("housing.xlsx")
install.packages("plyr")
library(plyr)
my_data1 <- numcolwise(sum)(my_data)
my_data2 <- t(my_data1)
colnames(my_data2)<-c("housing_market")
d <- my_data2
  year <- rownames(d)
  rownames(d) <- NULL
  data <- cbind(year,d)
data
```

A matrix: 8 x 2 of type chr

year	housing_market
2009	10756375
2010	11146943
2011	11519574
2012	11715862
2013	11994309
2014	12356841
2015	12443428
2016	12565782

```
# preparing labour productivity dataframe from[1]
my_data11 <- read_excel("labour productivity.xlsx")
my_data12 <- numcolwise(sum)(my_data11)
my_data21 <- t(my_data12)
colnames(my_data21)<-c("labour_productivity")
 d1 <- my_data21
 year <- rownames(d1)
 rownames(d1) <- NULL
 data1 <- cbind(year,d1)
data1
```

A matrix: 8 x 2 of type chr

year	labour_productivity
2009	1947.025
2010	1958.62
2011	1984.771
2012	2000
2013	2034.268
2014	2106.631
2015	2109.523
2016	2130.964

```
my_data111 <- read_excel("real gdp.xlsx")
my_data121 <- numcolwise(sum)(my_data111)
```

```
my_data211 <- t(my_data121)
colnames(my_data211)<-c("real_gdp")
 d11 <- my_data211
 year <- rownames(d11)
 rownames(d11) <- NULL
 data11 <- cbind(year,d11)
data11
```

A matrix: 8 x 2 of type chr

year	real_gdp
2009	1845.013
2010	1901.55
2011	1965.018
2012	2000
2013	2057.445
2014	2116.766
2015	2132.351
2016	2154.237

```
# preparing manufacturing activity dataframe from[1]
my_data1111 <- read_excel("manufacturing.xlsx")
my_data1112 <- t(my_data1111)
rownames(my_data1112) <- NULL
names(data3) <- lapply(data3[1, ], as.character)
data3 <- data3[-1,]
write.csv(data3,"File Name.csv", row.names = FALSE)
data3<- read.csv("File Name.csv", stringsAsFactors = FALSE)
names(data3)[names(data3) == "V1"] <- "year"
names(data3)[names(data3) == "V2"] <- "Manufacturing"
r1<-transform(data3, Manufacturing = as.numeric(Manufacturing))
x_new = aggregate(r1$Manufacturing, by=list(year=r1$year), FUN=sum)
x_new
```

A data.frame: 8 x 2

year	x
<int>	<dbl>
2009	317712.4
2010	331923.7
2011	348774.4
2012	360722.6
2013	361340.7
2014	372046.6
2015	386448.9
2016	189379.7

```
#transforming different column values of dataframes to numeric from chr

df1<-transform(df1, ngdp = as.numeric(ngdp))
data<-transform(data, year = as.numeric(as.character (year)))
data1<-transform(data1, year = as.numeric(as.character(year)))
data1<-transform(data1, labour_productivity = as.numeric(as.character(labour_productivity)))
data<-transform(data, year = as.numeric(as.character(year)))
data<-transform(data, housing_market = as.numeric(as.character(housing_market)))
data11<-transform(data11, real_gdp = as.numeric(as.character(real_gdp)))
df1
x_new
data
data1
data11
```

year	ngdp
<dbl>	<dbl>
2009	1.3700
2010	1.6100
2011	1.7888
2012	1.8243
2013	1.8428
2014	1.7993
2015	1.5598
2016	1.5358

A data.frame: 8 x 2

year	x
<int>	<dbl>
2009	317712.4
2010	331923.7
2011	348774.4
2012	360722.6
2013	381340.7
2014	372046.6
2015	388448.9
2016	189379.7

A data.frame: 8 x 2

year	housing market
<dbl>	<dbl>
2009	10758375
2010	11146943
2011	11519574
2012	11715882
2013	11994309
2014	12358841
2015	12443428
2016	12585782

A data.frame: 8 x 2

year	labour productivity
<dbl>	<dbl>
2009	1947.025
2010	1958.620
2011	1984.771
2012	2000.000
2013	2034.288
2014	2106.631
2015	2109.523
2016	2130.984

A data.frame: 8 x 2

year	real gdp
<dbl>	<dbl>
2009	1845.013
2010	1901.550
2011	1985.018
2012	2000.000
2013	2057.445
2014	2116.788
2015	2132.351
2016	2154.237

```
#merging dataframes 2 at a time
merged_df <- merge(df1, data1, by="year")
merged_df1 <- merge(x_new, data, by="year")
merged_df2 <- merge(merged_df1, merged_df, by="year")
merged_df3 <- merge(merged_df2, data11, by="year")
names(merged_df3)[names(merged_df3) == "x"] <- "Manufacturing"
merged_df3
```

A data.frame: 8 x 6

year <int>	Manufacturing <dbl>	housing_market <dbl>	ngdp <dbl>	labour_productivity <dbl>	real_gdp <dbl>
2009	317712.4	10756375	1.3700	1947.025	1845.013
2010	331923.7	11146943	1.6100	1958.620	1901.550
2011	348774.4	11519574	1.7886	1964.771	1965.018
2012	360722.6	11715862	1.8243	2000.000	2000.000
2013	361340.7	11994309	1.8426	2034.268	2057.445
2014	372046.6	12356841	1.7993	2106.631	2116.766
2015	386448.9	12443428	1.5596	2109.523	2132.351
2016	189379.7	12565782	1.5358	2130.964	2154.237

#applying correlation of explanatory variables to response variable nominal gdp and visualizing in r

```
#nominal gdp vs real gdp correlation
install.packages("ggpubr")
library("ggpubr")
ggscatter(merged_df3, x = "ngdp", y = "real_gdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "nominal", ylab = "real")
```

```
#nominal gdp vs labour_productivity correlation
library("ggpubr")
ggscatter(merged_df3, x = "labour_productivity", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "labour_productivity", ylab = "ngdp")
```

```
#nominal gdp vs housing market correlation
library("ggpubr")
ggscatter(merged_df3, x = "housing_market", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
```

```
        xlab = "housing_market", ylab = "ngdp")
```

```
#nominal gdp vs manufacturing activity correlation
library("ggpubr")
ggscatter(merged_df3, x = "Manufacturing", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "Manufacturing", ylab = "ngdp")
```

```
#nominal gdp vs inflation correlation
# putting inflation data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
inflation <- c(0.3,1.78,2.91,1.52,0.94, 1.91,1.13,1.43)
df2 <- data.frame(year, inflation)
#merging data frame with merged_df3
merged_df4 <- merge(merged_df3, df2, by="year")
# finding correlation between nominal gdp nd inflation
library("ggpubr")
ggscatter(merged_df4, x = "inflation", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "inflation", ylab = "ngdp")
```

```
#real interest rate vs inflation correlation
# putting real interest rate data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
interest_rate <- c(4.8,-.26,-.23,1.39, 1.76,1.01, 3.56,2.08)
df2 <- data.frame(year, interest_rate)
#merging data frame with merged_df3
merged_df4 <- merge(merged_df3, df2, by="year")
# finding correlation between nominal gdp nd interest_rate
library("ggpubr")
ggscatter(merged_df4, x = "interest_rate", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "interest_rate", ylab = "ngdp")
```

```
#nominal gdp vs income/wage growth correlation
```

```
# putting income/wage growth data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
income_growth <- c(-5.86,5.46,4.55,1.27,2.66,1.99,-1.58,0.85)
df2 <- data.frame(year, income_growth )
# #merging data frame with merged_df3
merged_df4 <- merge(merged_df4, df2, by="year")
merged_df4
# # finding correlation between nominal gdp nd income_growth
library("ggpubr")
ggscatter(merged_df4, x = "income_growth", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "income_growth ", ylab = "ngdp")
```

```
#nominal gdp vs investment correlation
# putting investment data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
investment <- c(1.53,1.84,2.14,2.71,3.64,3.56,3.95,2.25)
df2 <- data.frame(year, investment )
# #merging data frame with merged_df3
merged_df4 <- merge(merged_df4, df2, by="year")

# # finding correlation between nominal gdp nd investment
library("ggpubr")
ggscatter(merged_df4, x = "investment", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "investment", ylab = "ngdp")
```

```
#nominal gdp vs consumption
# putting consumption data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
consumption <- c(-1.11,2.42,1.29,.72,1.41,1.51,1.32,1.07)
df2 <- data.frame(year, consumption)
# #merging data frame with merged_df3
merged_df4 <- merge(merged_df4, df2, by="year")

# # finding correlation between nominal gdp nd consumption
library("ggpubr")
ggscatter(merged_df4, x = "consumption", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
```

```
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "consumption", ylab = "ngdp")
```

```
#nominal gdp vs stock_market
# putting stock_market data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
stock_market <- c(79.46,64.71,77.06,65.44,62.62,64.15,68.8,57.55)
df2 <- data.frame(year, stock_market)
# #merging data frame with merged_df3
merged_df4 <- merge(merged_df4, df2, by="year")

# # finding correlation between nominal gdp nd stock_market
library("ggpubr")
ggscatter(merged_df4, x = "stock_market", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "stock_market", ylab = "ngdp")
```

```
#nominal gdp vs tourist
# putting tourist data from [3] in a dataframe
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
tourist <- c(15.73,16.21,16.61,16.34,16.059,16.557,17.97,19.84)
df2 <- data.frame(year, tourist)
# #merging data frame with merged_df3
merged_df4 <- merge(merged_df4, df2, by="year")

# # finding correlation between nominal gdp nd tourist
library("ggpubr")
ggscatter(merged_df4, x = "tourist", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "tourist", ylab = "ngdp")
```

```
year <- c(2009,2010,2011,2012,2013,2014,2015,2016)
unemployment <- c(8.3,8.1,7.5,7.3,7.1,6.9,6.9,7)
df2 <- data.frame(year, unemployment)
# #merging data frame with merged_df3
merged_df4 <- merge(merged_df4, df2, by="year")

# # finding correlation between nominal gdp nd unemployment
```

```
library("ggpubr")
ggscatter(merged_df4, x = "unemployment", y = "ngdp",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
         xlab = "unemployment", ylab = "ngdp")
```

#Toronto Stock Exchange Trade Value Data

```
df_tse_val <- data.frame(year, ngdp)
df_all_stock <- read.csv("Stock_Market_Data.csv")
df_bond_rates <- read.csv("Bond Rates.csv")
df_corp_profs <- read.csv("Corporate Profits.csv")
df_all_stock %>% filter( ï..REF_DATE > "2008-12")
colnames(df_all_stock)
df_all_stock_2009 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
        grepl( "2009-",ï..REF_DATE))
tval_2009 <- sum(df_all_stock_2009$VALUE)
df_all_stock_2010 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
        grepl( "2010-",ï..REF_DATE))
tval_2010 <- sum(df_all_stock_2010$VALUE)
df_all_stock_2011 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
        grepl( "2011-",ï..REF_DATE))
tval_2011 <- sum(df_all_stock_2011$VALUE)
df_all_stock_2012 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
        grepl( "2012-",ï..REF_DATE))
tval_2012 <- sum(df_all_stock_2012$VALUE)
df_all_stock_2013 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
        grepl( "2013-",ï..REF_DATE))
tval_2013 <- sum(df_all_stock_2013$VALUE)
df_all_stock_2014 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
        grepl( "2014-",ï..REF_DATE))
tval_2014 <- sum(df_all_stock_2014$VALUE)
df_all_stock_2015 <- df_all_stock %>%
```

```
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
         grepl( "2015-",ï..REF_DATE))
tval_2015 <- sum(df_all_stock_2015$VALUE)
df_all_stock_2016 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, value of shares traded",Stock.market.statistics),
         grepl( "2016-",ï..REF_DATE))
tval_2016 <- sum(df_all_stock_2016$VALUE)
tval <- c[tval_2009,
          tval_2010,
          tval_2011,
          tval_2012,
          tval_2013,
          tval_2014,
          tval_2015,
          tval_2016]
df_tse <- data.frame(year, ngdp, tval)
df_tse

#nominal gdp vs value of shares traded
ggscatter(df_tse, x = "ngdp", y = "tval",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Toronto Stock Exchange Value traded", ylab = "real")

#================================================================================
#Toronto Stock Exchange Volume Datadf_all_stock_v2009 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2009-",ï..REF_DATE))
tvol_2009 <- sum(df_all_stock_v2009$VALUE)
df_all_stock_v2010 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2010-",ï..REF_DATE))
tvol_2010 <- sum(df_all_stock_v2010$VALUE)
df_all_stock_v2011 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2011-",ï..REF_DATE))
tvol_2011 <- sum(df_all_stock_v2011$VALUE)
df_all_stock_v2012 <- df_all_stock %>%
```

```
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2012-",ï..REF_DATE))
tvol_2012 <- sum(df_all_stock_v2012$VALUE)
df_all_stock_v2013 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2013-",ï..REF_DATE))
tvol_2013 <- sum(df_all_stock_v2013$VALUE)
df_all_stock_v2014 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2014-",ï..REF_DATE))
tvol_2014 <- sum(df_all_stock_v2014$VALUE)
df_all_stock_v2015 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
         grepl( "2015-",ï..REF_DATE))
tvol_2015 <- sum(df_all_stock_v2015$VALUE)
df_all_stock_v2016 <- df_all_stock %>%
  filter(grepl( "Toronto Stock Exchange, volume of shares traded",Stock.market.statistics),
grepl( "2016-",ï..REF_DATE))
tvol_2016 <- sum(df_all_stock_v2016$VALUE)
tvol <- c(tvol_2009,
          tvol_2010,
          tvol_2011,
          tvol_2012,
          tvol_2013,
          tvol_2014,
          tvol_2015,
          tvol_2016)
df_tse <- data.frame(year, ngdp, tval,tvol)
df_tse
#nominal gdp vs vol of shares traded
ggscatter(df_tse, x = "ngdp", y = "tvol",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Toronto Stock Exchange Volume Traded", ylab = "real")

#=============================================================================
# Selected Government of Canada benchmark bond yields: 10 years

colnames(df_bond_rates)
```

```
df_bond <- df_bond_rates %>%
  filter(grepl( "Selected Government of Canada benchmark bond yields: 10 years",Rates),
       grepl( "2009-|2010-|2011-|2012-|2013-|2014-|2015-|2016-",ï..REF_DATE))
df_bond
df_bond_2009 <- df_bond %>% filter(grepl("2009",ï..REF_DATE))
df_bond_2009 <- sum(df_bond_2009$VALUE)
bon_val_2009 <- (df_bond_2009/12)
bon_val_2009
df_bond_2010 <- df_bond %>% filter(grepl("2010",ï..REF_DATE))
df_bond_2010 <- sum(df_bond_2010$VALUE)
bon_val_2010 <- (df_bond_2010/12)
df_bond_2011 <- df_bond %>% filter(grepl("2011",ï..REF_DATE))
df_bond_2011 <- sum(df_bond_2011$VALUE)
bon_val_2011 <- (df_bond_2011/12)



df_bond_2012 <- df_bond %>% filter(grepl("2012",ï..REF_DATE))
df_bond_2012 <- sum(df_bond_2012$VALUE)
bon_val_2012 <- (df_bond_2012/12)
df_bond_2013 <- df_bond %>% filter(grepl("2013",ï..REF_DATE))
df_bond_2013 <- sum(df_bond_2013$VALUE)
bon_val_2013 <- (df_bond_2013/12)
df_bond_2014 <- df_bond %>% filter(grepl("2014",ï..REF_DATE))
df_bond_2014 <- sum(df_bond_2014$VALUE)
bon_val_2014 <- (df_bond_2014/12)
df_bond_2015 <- df_bond %>% filter(grepl("2015",ï..REF_DATE))
df_bond_2015 <- sum(df_bond_2015$VALUE)
bon_val_2015 <- (df_bond_2015/12)
df_bond_2016 <- df_bond %>% filter(grepl("2016",ï..REF_DATE))
df_bond_2016 <- sum(df_bond_2016$VALUE)
bon_val_2016 <- (df_bond_2016/12)
bond_10yr <- c(bon_val_2009,
           bon_val_2010,
           bon_val_2011,
           bon_val_2012,
           bon_val_2013,
           bon_val_2014,
           bon_val_2015,
```

```
        bon_val_2016)
bond_10yr
df_tse <- data.frame(year, ngdp, tval, tvol, bond_10yr)
df_tse
#nominal gdp vs vol of shares traded
ggscatter(df_tse, x = "ngdp", y = "bond_10yr",
      add = "reg.line", conf.int = TRUE,
      cor.coef = TRUE, cor.method = "pearson",
      xlab = "Selected Government of Canada benchmark bond yields: 10 years", ylab = "real")

#=============================================================================
# EUR_CAD Conversion Rates
eur_cad <- c(1.59,1.37,1.38,1.28,1.37,1.47,1.42,1.47)

df_tse <- data.frame(year, ngdp, tval, tvol, bond_10yr,eur_cad)
df_tse

ggscatter(df_tse, x = "ngdp", y = "eur_cad",
      add = "reg.line", conf.int = TRUE,
      cor.coef = TRUE, cor.method = "pearson",
      xlab = "EUR to CAD Exchange Rates", ylab = "real")

#=============================================================================
#USD To CAD Conversion Rates
usd_cad <- c(1.141433,1.030533,0.989323,0.999997,1.030084,1.104347,1.279163,1.325521)

df_tse <- data.frame(year, ngdp, tval, tvol, bond_10yr,eur_cad, usd_cad)
df_tse

ggscatter(df_tse, x = "ngdp", y = "usd_cad",
      add = "reg.line", conf.int = TRUE,
      cor.coef = TRUE, cor.method = "pearson",
      xlab = "USD to CAD Exchange Rates", ylab = "real")

#=============================================================================
#Labour Productivity Rates

lab_prod <- c(53.8,54.6,55.6,55.8,56.7,58.1,58.1,58.3)
```

```
df_tse <- data.frame(year, ngdp, tval, tvol, bond_10yr,eur_cad, usd_cad, lab_prod)
df_tse

ggscatter(df_tse, x = "ngdp", y = "lab_prod",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Labour Productivity", ylab = "real")
#================================================================================
#Employment Rate
emp_level <- c(16.73,16.96,17.22,17.44,17.69,17.8,17.95,18.08)
df_tse <- data.frame(year, ngdp, tval, tvol, bond_10yr,eur_cad, usd_cad, lab_prod, emp_level)
df_tse

ggscatter(df_tse, x = "ngdp", y = "emp_level",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Employment Rate", ylab = "real")

#================================================================================
#Ease of Doing Business
eob <- c(8,12,13,17,19,13,20,22)
df_tse <- data.frame(year, ngdp, tval, tvol, bond_10yr,eur_cad, usd_cad, lab_prod, emp_level,eob)
df_tse

ggscatter(df_tse, x = "ngdp", y = "eob",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Ease of Doing Business", ylab = "real")
```