Alternative Final Exercise

Loading Libraries

rm(list=ls())  
library(haven)  
library(readxl)  
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(ggplot2)  
library(plm)

## Loading required package: Formula

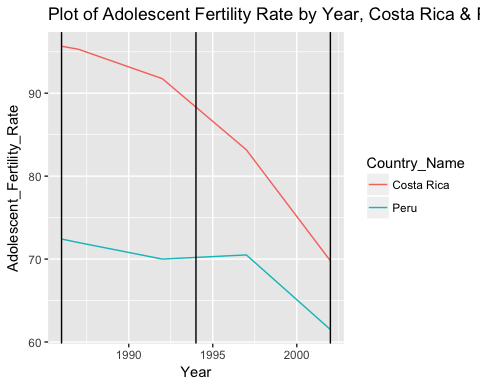
##   
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':  
##   
## between, lag, lead

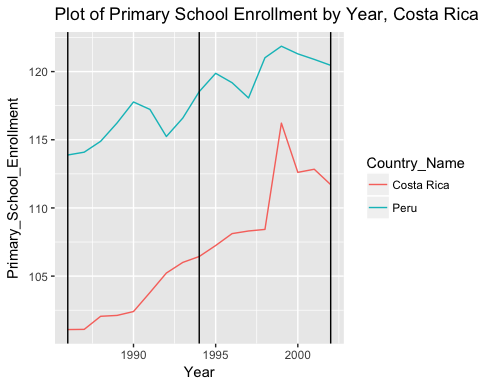
Loading Data

# Load in WDI Data  
wdi.data <- read\_dta("~/desktop/stats/final\_exercise/wdi.dta")  
  
# LOAD WDI COUNTRIES  
wdi.countries <- read\_excel("~/desktop/stats/final\_exercise/Country.xlsx")

# Selected Attributes  
wdi.selected <- select(wdi.data, countryname, countrycode, year, sp\_ado\_tfrt, se\_sec\_enrr, se\_sec\_enrr\_fe, se\_sec\_enrr\_ma, ny\_gdp\_pcap\_cd, si\_pov\_gini, sp\_pop\_totl\_fe\_zs, se\_com\_durs, sp\_dyn\_conu\_zs, sp\_uwt\_tfrt, sp\_pop\_totl, sh\_sta\_mmrt\_ne, sp\_rur\_totl\_zs, sl\_tlf\_0714\_fe\_zs, sl\_tlf\_acti\_1524\_fe\_ne\_zs, se\_prm\_enrr)  
  
# Rename Columns  
names(wdi.selected) <- c("Country\_Name", "Country Code", "Year", "Adolescent\_Fertility\_Rate","Secondary\_School\_Enrollment", "Female\_Secondary\_School\_Enrollment", "Male\_Secondary\_School\_Enrollment", "GDP\_Per\_Capita", "GINI", "Percent\_Female", "Compulsory\_Education", "Contraceptive Prevalence", "Unmet Need for Contraception", "Population", "Maternal\_Mortality\_Rate", "Percent\_Rural\_Population", "Female\_Child\_Employment", "Female\_Labor\_Force\_Participation", "Primary\_School\_Enrollment")  
  
wdi.did <- filter(wdi.selected, (Country\_Name =="Costa Rica" | Country\_Name=="Peru") & (Year >= 1986 & Year <= 2002))  
  
wdi.more <- filter(wdi.selected, (Country\_Name =="Costa Rica" | Country\_Name=="Peru" | Country\_Name=="Panama" | Country\_Name=="Guatemala" | Country\_Name=="Honduras" | Country\_Name=="Dominican Republic" | Country\_Name=="Ecuador" | Country\_Name=="Venezuela, RB" | Country\_Name=="Colombia" | Country\_Name=="Paraguay") & (Year >= 1986 & Year <= 2002))  
  
ggplot(data=wdi.did)+   
 geom\_line(aes(`Year`, `Adolescent\_Fertility\_Rate`, group=`Country\_Name`, colour=`Country\_Name`),) +   
 ggtitle("Plot of Adolescent Fertility Rate by Year, Costa Rica & Peru") +  
 geom\_vline(xintercept = 1994) +  
 geom\_vline(xintercept = 1986) +  
 geom\_vline(xintercept = 2002)

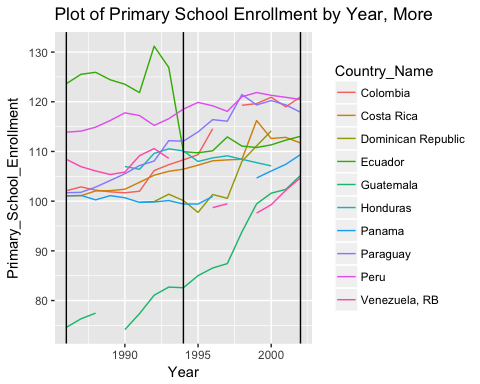
 Primary School Enrollment Regressions

ggplot(data=wdi.did)+   
 geom\_line(aes(`Year`, `Primary\_School\_Enrollment`, group=`Country\_Name`, colour=`Country\_Name`),) +  
 ggtitle("Plot of Primary School Enrollment by Year, Costa Rica & Peru") +  
 geom\_vline(xintercept = 1994) +  
 geom\_vline(xintercept = 1986) +  
 geom\_vline(xintercept = 2002)



ggplot(data=wdi.more)+   
 geom\_line(aes(`Year`, `Primary\_School\_Enrollment`, group=`Country\_Name`, colour=`Country\_Name`),) +  
 ggtitle("Plot of Primary School Enrollment by Year, More") +  
 geom\_vline(xintercept = 1994) +  
 geom\_vline(xintercept = 1986) +  
 geom\_vline(xintercept = 2002)

## Warning: Removed 2 rows containing missing values (geom\_path).

 Set up Diff-in-Diff Variables

# Set "intervention"" variable for all observations after 1994  
wdi.did$intervention <- ifelse(wdi.did$Year > 1994, 1, 0)  
# Set "treated" for all treated countries (i.e. Costa Rica)  
wdi.did$treated <- ifelse(wdi.did$Country\_Name == "Costa Rica", 1, 0)  
# Set up the interaction variable "did" to be intervention \* treated  
# In this case set 1 for observations of Costa Rica after 1994 when the intervention occured  
wdi.did$did <- wdi.did$intervention \* wdi.did$treated

Primary School Regressions Regress Adolescent Fertility Rate on Primary School Enrollment

# Regress Adolescent Fertility Rate on Primary School Enrollment  
primarySchoolEnrollReg <- lm(data=wdi.did, Adolescent\_Fertility\_Rate ~ Primary\_School\_Enrollment + GDP\_Per\_Capita)  
  
summary(primarySchoolEnrollReg)

##   
## Call:  
## lm(formula = Adolescent\_Fertility\_Rate ~ Primary\_School\_Enrollment +   
## GDP\_Per\_Capita, data = wdi.did)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.6404 -1.8750 0.5987 1.6743 6.1687   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.489e+02 9.780e+00 25.454 <2e-16 \*\*\*  
## Primary\_School\_Enrollment -1.519e+00 8.247e-02 -18.419 <2e-16 \*\*\*  
## GDP\_Per\_Capita -2.006e-04 6.123e-04 -0.328 0.745   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.086 on 31 degrees of freedom  
## Multiple R-squared: 0.9216, Adjusted R-squared: 0.9166   
## F-statistic: 182.2 on 2 and 31 DF, p-value: < 2.2e-16

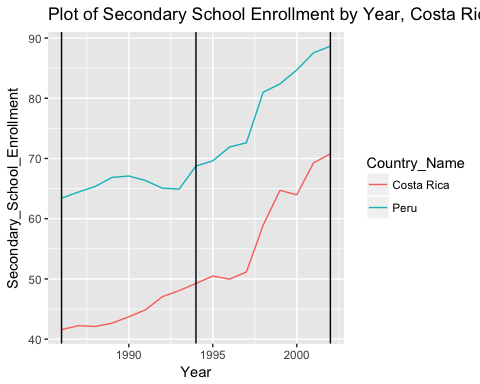
Diff-in-Diff regression to show impact of intervention on Primary School Enrollment First Stage Regression (see bottom for second stage regresssion)

# Diff-in-Diff regression to show impact of intervention on Primary School Enrollment  
# When you control for GDP\_Per\_Capita the significance disappears....  
primaryDidReg <- lm(Primary\_School\_Enrollment ~ treated + intervention + did, data = wdi.did)  
  
summary(primaryDidReg)

##   
## Call:  
## lm(formula = Primary\_School\_Enrollment ~ treated + intervention +   
## did, data = wdi.did)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4404 -1.7962 0.1549 1.4410 5.5363   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 116.0494 0.7096 163.535 < 2e-16 \*\*\*  
## treated -12.6925 1.0036 -12.647 1.48e-13 \*\*\*  
## intervention 4.2818 1.0345 4.139 0.00026 \*\*\*  
## did 3.0451 1.4629 2.081 0.04602 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.129 on 30 degrees of freedom  
## Multiple R-squared: 0.9105, Adjusted R-squared: 0.9015   
## F-statistic: 101.7 on 3 and 30 DF, p-value: 8.165e-16

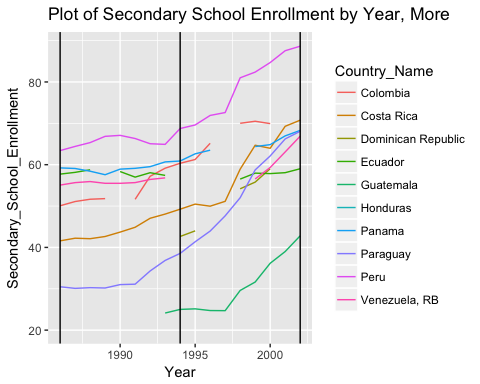
Secondary School Enrollment Charts

ggplot(data=wdi.did)+   
 geom\_line(aes(`Year`, `Secondary\_School\_Enrollment`, group=`Country\_Name`, colour=`Country\_Name`),) +  
 ggtitle("Plot of Secondary School Enrollment by Year, Costa Rica & Peru") +  
 geom\_vline(xintercept = 1994) +  
 geom\_vline(xintercept = 1986) +  
 geom\_vline(xintercept = 2002)



ggplot(data=wdi.more)+   
 geom\_line(aes(`Year`, `Secondary\_School\_Enrollment`, group=`Country\_Name`, colour=`Country\_Name`),) +  
 ggtitle("Plot of Secondary School Enrollment by Year, More") +  
 geom\_vline(xintercept = 1994) +  
 geom\_vline(xintercept = 1986) +  
 geom\_vline(xintercept = 2002)

## Warning: Removed 9 rows containing missing values (geom\_path).

 Secondary School Enrollment Regressions Regress Adolescent Fertility Rate on Secondary School Enrollment

secondarySchoolEnrollReg <- lm(data=wdi.did, Adolescent\_Fertility\_Rate ~ Secondary\_School\_Enrollment + GDP\_Per\_Capita)  
  
summary(secondarySchoolEnrollReg)

##   
## Call:  
## lm(formula = Adolescent\_Fertility\_Rate ~ Secondary\_School\_Enrollment +   
## GDP\_Per\_Capita, data = wdi.did)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7675 -2.4310 -0.2971 2.6618 4.8119   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.201e+02 2.971e+00 40.42 <2e-16 \*\*\*  
## Secondary\_School\_Enrollment -7.226e-01 3.819e-02 -18.92 <2e-16 \*\*\*  
## GDP\_Per\_Capita 1.043e-03 5.825e-04 1.79 0.0833 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.011 on 31 degrees of freedom  
## Multiple R-squared: 0.9254, Adjusted R-squared: 0.9206   
## F-statistic: 192.2 on 2 and 31 DF, p-value: < 2.2e-16

First State Regression Diff-in-Diff regression to show impact of intervention on the Secondary School Enrollment

# Diff-in-Diff regression to show impact of intervention on Secondary School Enrollment  
# When you control for GDP\_Per\_Capita the significance disappears....  
secondaryDidReg <- lm(Secondary\_School\_Enrollment ~ treated + intervention + did, data = wdi.did)  
  
summary(secondaryDidReg)

##   
## Call:  
## lm(formula = Secondary\_School\_Enrollment ~ treated + intervention +   
## did, data = wdi.did)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.1967 -2.3852 -0.0971 3.3242 10.8572   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 65.796 1.910 34.442 < 2e-16 \*\*\*  
## treated -21.188 2.702 -7.842 9.44e-09 \*\*\*  
## intervention 14.005 2.785 5.029 2.15e-05 \*\*\*  
## did 1.289 3.938 0.327 0.746   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.731 on 30 degrees of freedom  
## Multiple R-squared: 0.8462, Adjusted R-squared: 0.8308   
## F-statistic: 55.02 on 3 and 30 DF, p-value: 2.639e-12

Second State Regression Diff-in-Diff regression to show the impact of the intervention on the Adolescent Fertility Rate

fertilityDidReg <- lm(Adolescent\_Fertility\_Rate ~ treated + intervention + did + GDP\_Per\_Capita, data = wdi.did)  
  
summary(fertilityDidReg)

##   
## Call:  
## lm(formula = Adolescent\_Fertility\_Rate ~ treated + intervention +   
## did + GDP\_Per\_Capita, data = wdi.did)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.3966 -1.1767 0.1593 1.4945 5.6418   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 78.655932 2.366961 33.231 < 2e-16 \*\*\*  
## treated 28.717862 2.396757 11.982 9.4e-13 \*\*\*  
## intervention 1.419999 2.094208 0.678 0.50311   
## did -7.814527 2.203090 -3.547 0.00135 \*\*   
## GDP\_Per\_Capita -0.006104 0.001695 -3.602 0.00116 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.067 on 29 degrees of freedom  
## Multiple R-squared: 0.9275, Adjusted R-squared: 0.9175   
## F-statistic: 92.8 on 4 and 29 DF, p-value: 4.279e-16