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

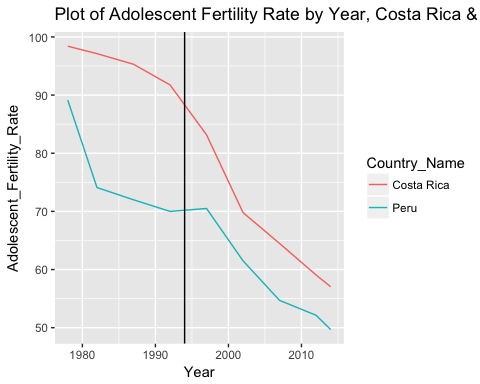
##   
## Attaching package: 'plm'

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

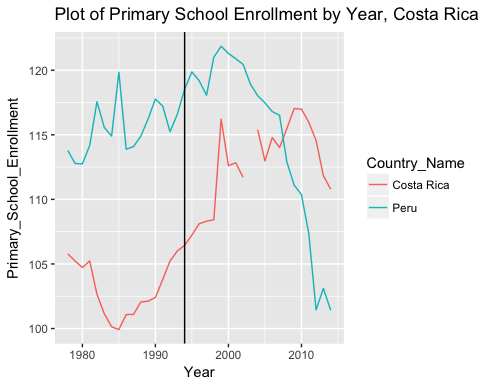
Loading Data

# 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/FinalExercise/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 >= 1978 & Year <= 2014))  
  
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 >= 1978 & Year <= 2014))  
  
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)

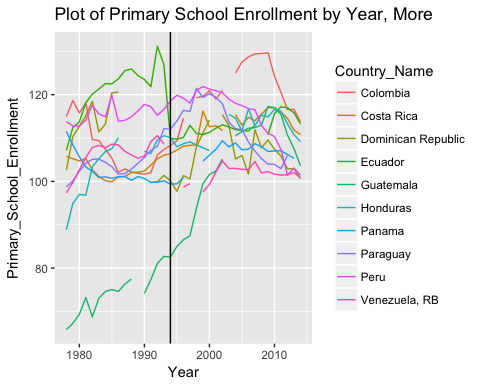
 Primary School Enrollment Regressions

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)



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)

## Warning: Removed 3 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   
## -26.822 -4.088 2.584 5.359 12.164   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.334e+02 1.861e+01 12.546 < 2e-16 \*\*\*  
## Primary\_School\_Enrollment -1.330e+00 1.653e-01 -8.046 1.51e-11 \*\*\*  
## GDP\_Per\_Capita -3.399e-03 4.355e-04 -7.803 4.21e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.924 on 70 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.6357, Adjusted R-squared: 0.6253   
## F-statistic: 61.08 on 2 and 70 DF, p-value: 4.474e-16

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

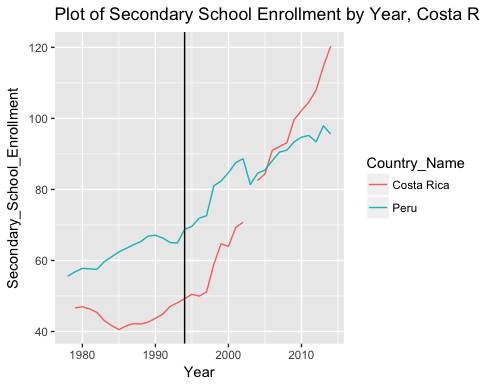
# Diff-in-Diff regression to show impact of intervention on Primary School Enrollment  
# When you control for GDP\_Per\_Capita the significance disappears....  
primaryDidReg <- lm(Primary\_School\_Enrollment ~ treated + intervention + did, data = wdi.did)  
  
summary(primaryDidReg)

##   
## Call:  
## lm(formula = Primary\_School\_Enrollment ~ treated + intervention +   
## did, data = wdi.did)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.4830 -2.0137 0.5595 2.5759 6.9556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 115.6352 1.0135 114.092 < 2e-16 \*\*\*  
## treated -12.3983 1.4333 -8.650 1.30e-12 \*\*\*  
## intervention -0.7301 1.3785 -0.530 0.598   
## did 10.4039 1.9613 5.305 1.29e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.179 on 69 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.587, Adjusted R-squared: 0.5691   
## F-statistic: 32.69 on 3 and 69 DF, p-value: 2.912e-13

Secondary School Enrollment Charts

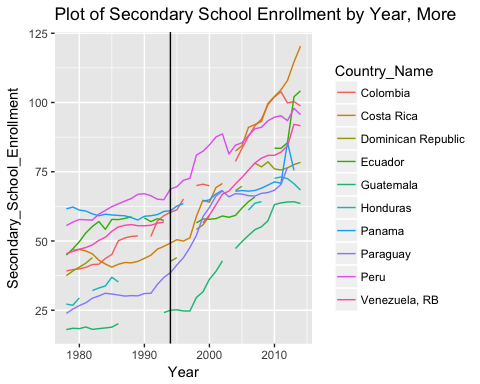
ggplot(data=wdi.did)+   
 geom\_line(aes(`Year`, `Secondary\_School\_Enrollment`, group=`Country\_Name`, colour=`Country\_Name`),) +  
 ggtitle("Plot of Secondary School Enrollment by Year, Costa Rica & Peru") +  
 geom\_vline(xintercept = 1994)

## Warning: Removed 1 rows containing missing values (geom\_path).



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)

## Warning: Removed 4 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   
## -9.922 -3.019 -0.684 2.509 9.209   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 124.667589 2.145403 58.109 < 2e-16 \*\*\*  
## Secondary\_School\_Enrollment -0.824310 0.039413 -20.915 < 2e-16 \*\*\*  
## GDP\_Per\_Capita 0.002119 0.000338 6.271 2.71e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.522 on 69 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.9039, Adjusted R-squared: 0.9012   
## F-statistic: 324.6 on 2 and 69 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   
## -32.748 -4.232 0.933 4.596 37.610   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 62.387 3.058 20.401 < 2e-16 \*\*\*  
## treated -17.879 4.392 -4.071 0.000124 \*\*\*  
## intervention 24.099 4.159 5.794 1.94e-07 \*\*\*  
## did 14.109 5.967 2.364 0.020920 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.61 on 68 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.6477, Adjusted R-squared: 0.6321   
## F-statistic: 41.67 on 3 and 68 DF, p-value: 2.126e-15

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   
## -6.2321 -3.0938 -0.5528 2.1182 13.4025   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 78.2097081 1.0843132 72.128 < 2e-16 \*\*\*  
## treated 23.2383224 1.4762253 15.742 < 2e-16 \*\*\*  
## intervention -7.2389026 1.5758604 -4.594 1.91e-05 \*\*\*  
## did -5.9453565 2.0240397 -2.937 0.0045 \*\*   
## GDP\_Per\_Capita -0.0032833 0.0003094 -10.611 3.83e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.232 on 69 degrees of freedom  
## Multiple R-squared: 0.9194, Adjusted R-squared: 0.9147   
## F-statistic: 196.7 on 4 and 69 DF, p-value: < 2.2e-16