

Maternal Mortality Ratio

Linear Regression
Exploration

Samantha Jackson

Agenda

Results

Data Collection

EDA

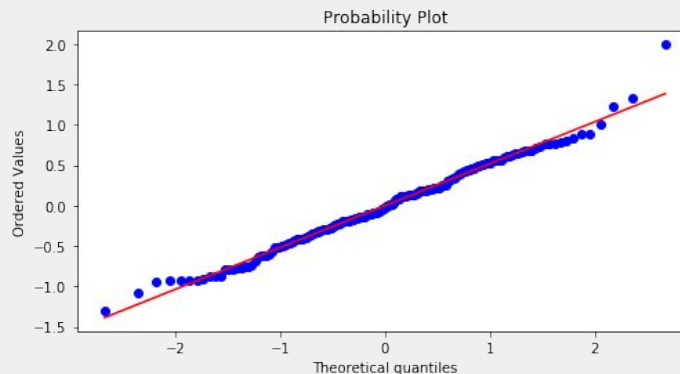
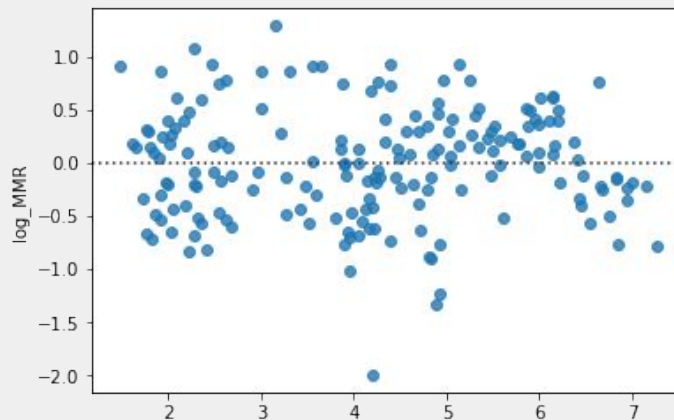
Conclusions



Results

Dep. Variable:	log_MMR	R-squared:	0.902
Model:	OLS	Adj. R-squared:	0.897
Method:	Least Squares	F-statistic:	177.3
Date:	Sun, 21 Apr 2019	Prob (F-statistic):	1.82e-82
Time:	22:10:29	Log-Likelihood:	-138.57
No. Observations:	183	AIC:	297.1
Df Residuals:	173	BIC:	329.2
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	9.8799	0.622	15.892	0.000	8.653	11.107
Teen_fertility	0.0045	0.002	2.707	0.007	0.001	0.008
Immunization_measles	-0.0178	0.004	-4.450	0.000	-0.026	-0.010
Internet_use	-0.0166	0.004	-4.465	0.000	-0.024	-0.009
log_GDP	-0.3977	0.058	-6.848	0.000	-0.512	-0.283
North_America	-0.4318	0.145	-2.983	0.003	-0.718	-0.146
Oceania	-0.7289	0.202	-3.605	0.000	-1.128	-0.330
Asia	-0.6865	0.127	-5.386	0.000	-0.938	-0.435
Europe	-2.4308	0.319	-7.632	0.000	-3.059	-1.802
IUxEurope	0.0199	0.005	3.980	0.000	0.010	0.030



Testing:
Linear relationship
between world
development
indicators and
Maternal Mortality
Ratio

After a number of
iterations, the best
results, with an
R-squared of .902,
came from using a
log transformed
dependant variable.

Data

- Pulled data from The World Bank Databank, world development indicators database
- Data was output to a CSV
- Initial dataset included 69 features/development indicators for 216 countries over 40 years
- Data points were by country & feature, with years as columns. I needed to transform my data to be by country & year, with features as columns.

Variable Definitions

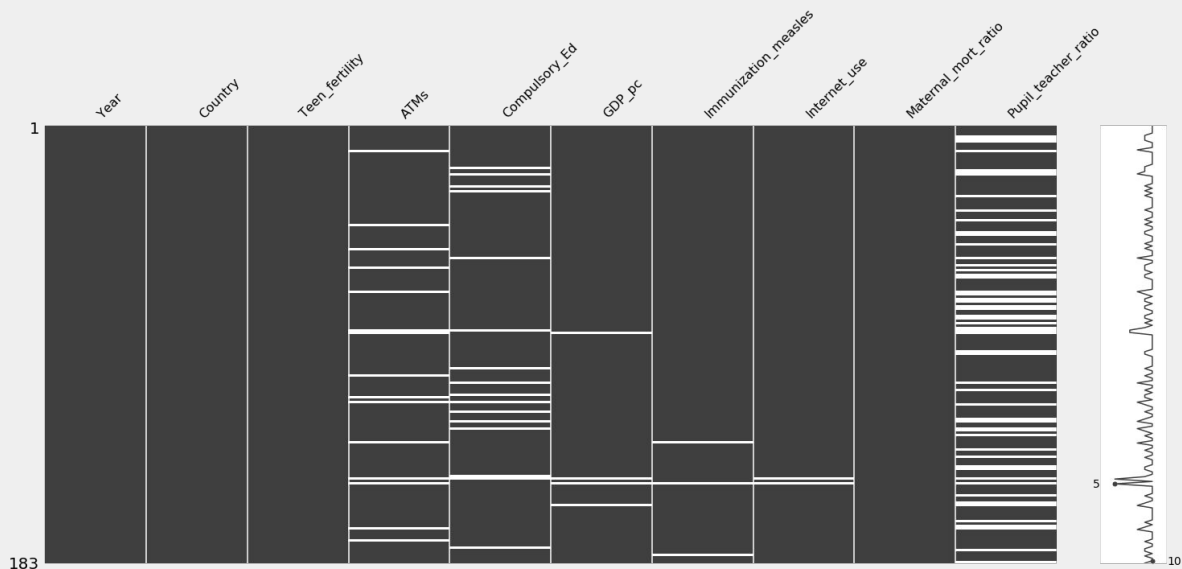
- **Maternal Mortality Ratio:** Maternal deaths per 100K live births
- **Teen Fertility:** Births per 100K women ages 15-19
- **Immunization Measles:** % of children ages 12-23 months immunized for measles
- **Internet Use:** % of population using the internet
- **GDP_pc:** Gross Domestic Product per capita
- **Continent**



THE WORLD BANK
IBRD • IDA

EDA - Missing Data

Narrowed down data using a process of checking the number of datapoints available for each feature & checking correlation matrices. My goal was to minimize multicollinearity while maximizing correlation with the target variable. I then chose a year with the most data points available - 2010.



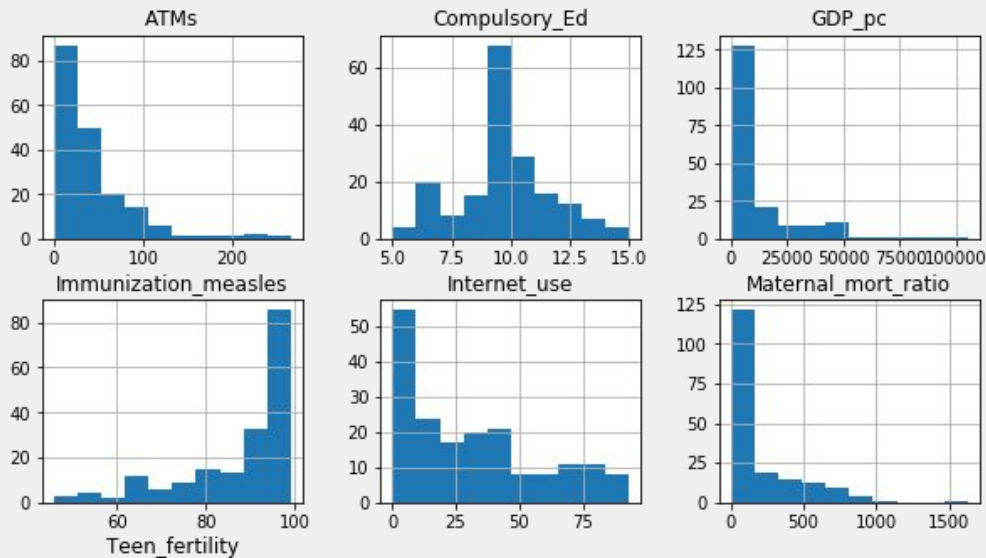
Imputation: Didn't want to take full dataset averages for all categories, so I scraped Region and Continent for each country.

- **ATMs:** Mean of Region
- **Compulsory_Ed:** Mean of all countries
- **GDP_pc:** Median of Region
- **Immunizations:** Mean of Region
- **Internet Use:** Mean of Region
- **Pupil_teacher_ratio:** Dropped from Dataset

Variable Definitions

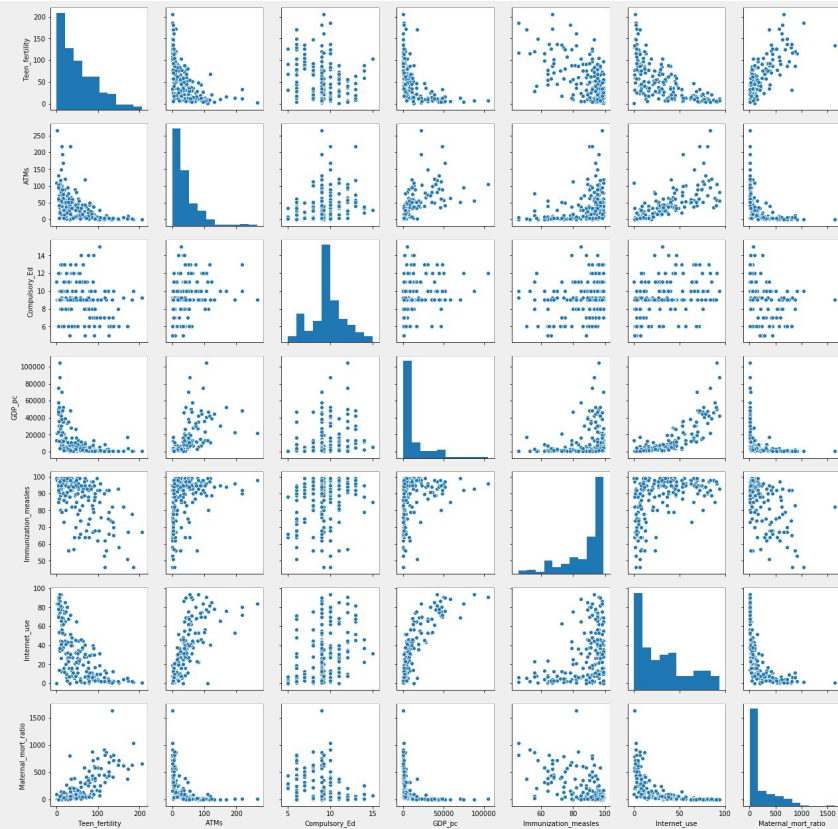
- **Maternal Mortality Ratio:** Maternal deaths per 100K live births
- **Teen Fertility:** Births per 100K women ages 15-19
- **Immunization Measles:** % of children ages 12-23 months immunized for measles
- **Internet Use:** % of population using the internet
- **GDP_pc:** Gross Domestic Product per capita
- **Continent**

EDA - Feature Analysis



Variable Definitions

- **Maternal Mortality Ratio:** Maternal deaths per 100K live births
- **Teen Fertility:** Births per 100K women ages 15-19
- **Immunization Measles:** % of children ages 12-23 months immunized for measles
- **Internet Use:** % of population using the internet
- **GDP_pc:** Gross Domestic Product per capita
- **Continent**



EDA – Initial Model

Dep. Variable:	Maternal_mort_ratio	R-squared:	0.652
Model:	OLS	Adj. R-squared:	0.640
Method:	Least Squares	F-statistic:	55.04
Date:	Thu, 18 Apr 2019	Prob (F-statistic):	7.23e-38
Time:	10:36:50	Log-Likelihood:	-1184.0
No. Observations:	183	AIC:	2382.
Df Residuals:	176	BIC:	2405.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	493.2773	123.282	4.001	0.000	249.977	736.578
Teen_fertility	3.1526	0.386	8.162	0.000	2.390	3.915
ATMs	-0.1796	0.393	-0.457	0.648	-0.955	0.596
Compulsory_Ed	-8.2090	6.414	-1.280	0.202	-20.867	4.449
GDP_pc	0.0025	0.001	2.099	0.037	0.000	0.005
Immunization_measles	-3.7410	1.160	-3.225	0.002	-6.030	-1.452
Internet_use	-2.8448	0.965	-2.949	0.004	-4.748	-0.941

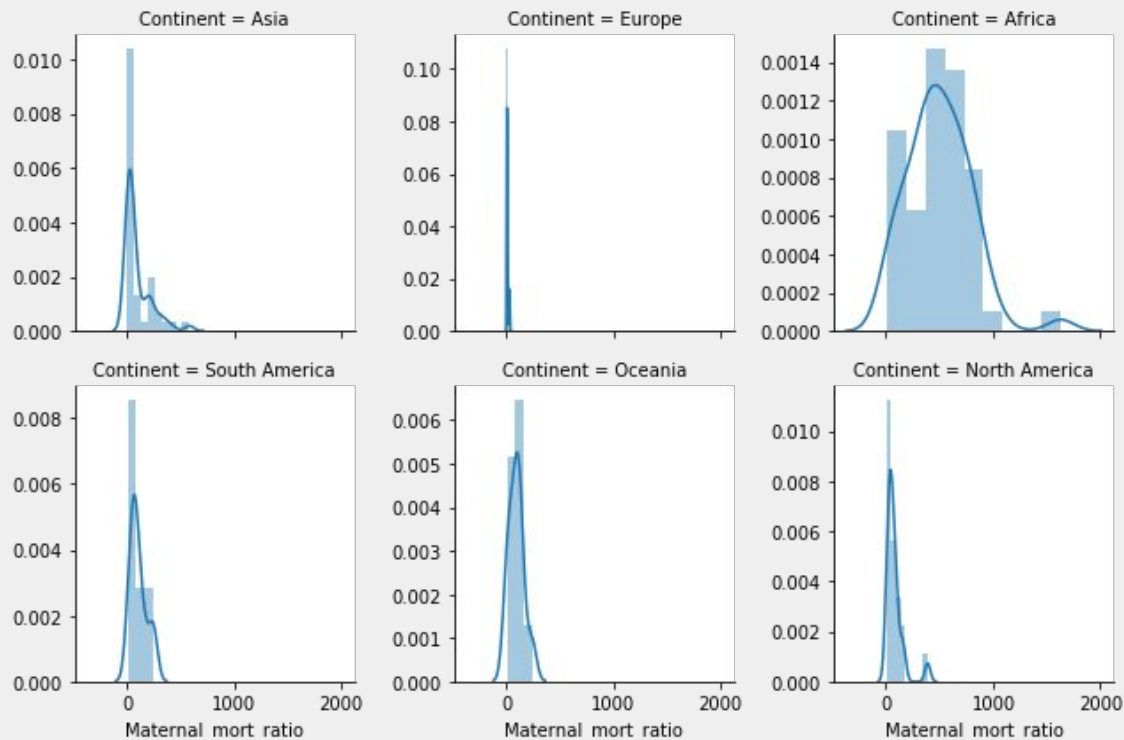
First Model - using non-transformed features & non-transformed target

Dep. Variable:	log_MMR	R-squared:	0.835
Model:	OLS	Adj. R-squared:	0.830
Method:	Least Squares	F-statistic:	148.6
Date:	Thu, 18 Apr 2019	Prob (F-statistic):	3.56e-66
Time:	10:42:04	Log-Likelihood:	-186.35
No. Observations:	183	AIC:	386.7
Df Residuals:	176	BIC:	409.2
Df Model:	6		
Covariance Type:	nonrobust		

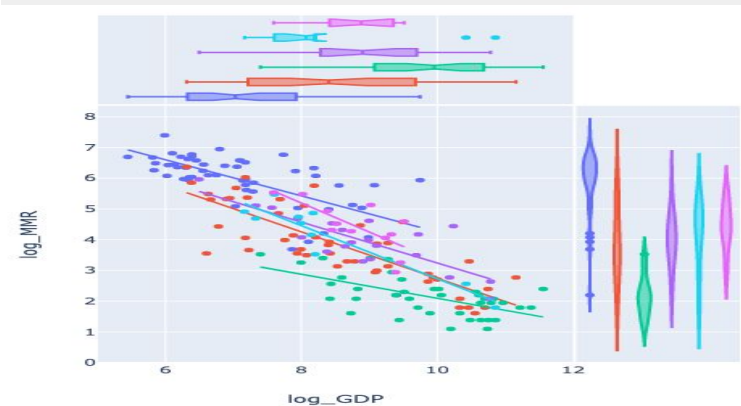
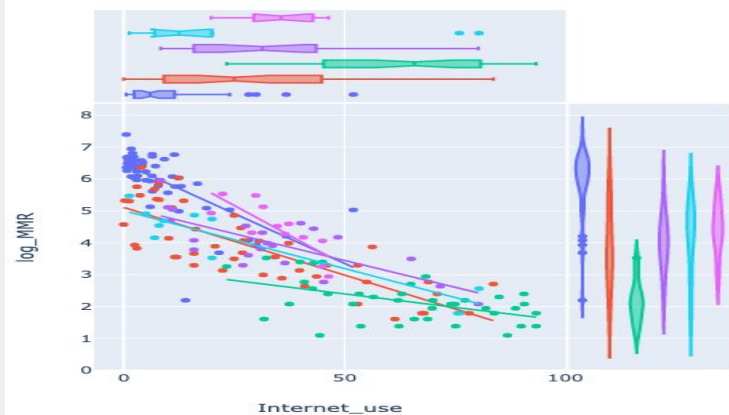
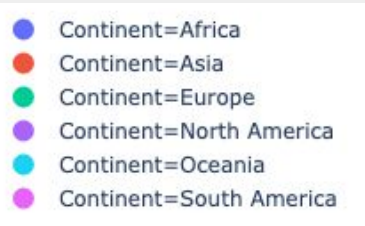
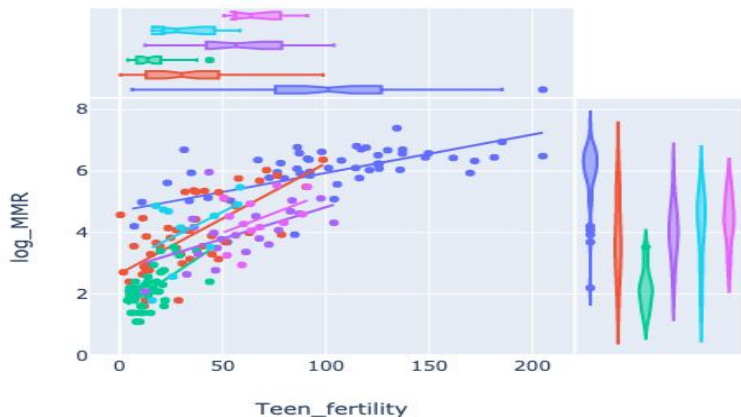
	coef	std err	t	P> t	[0.025	0.975]
Intercept	5.8645	0.529	11.093	0.000	4.821	6.908
Teen_fertility	0.0133	0.002	8.030	0.000	0.010	0.017
ATMs	-0.0021	0.002	-1.239	0.217	-0.005	0.001
Compulsory_Ed	-0.0304	0.028	-1.107	0.270	-0.085	0.024
GDP_pc	2.828e-06	5.02e-06	0.563	0.574	-7.08e-06	1.27e-05
Immunization_measles	-0.0116	0.005	-2.341	0.020	-0.021	-0.002
Internet_use	-0.0340	0.004	-8.212	0.000	-0.042	-0.026

Second Model - log-transformed target only

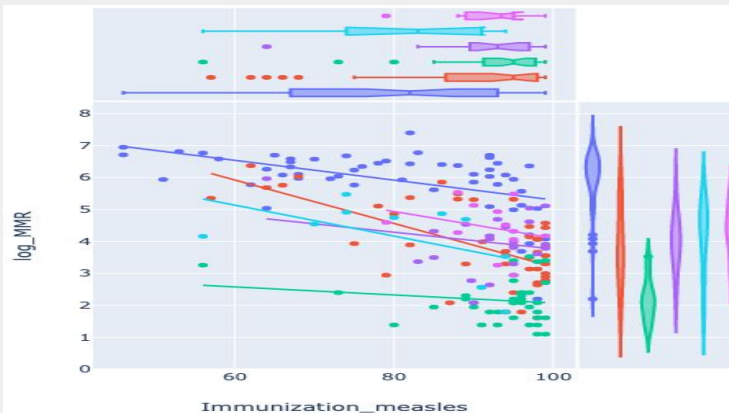
EDA – Target distribution by Continent



EDA - Interactions



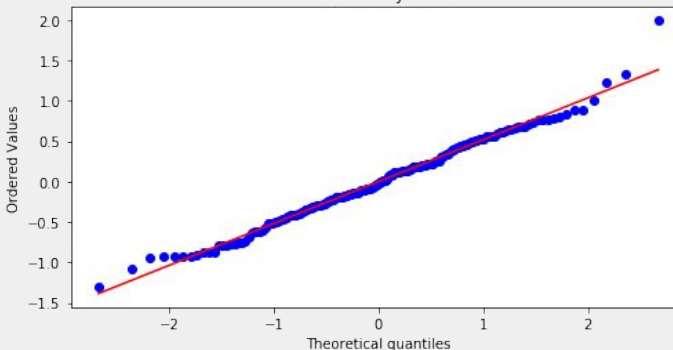
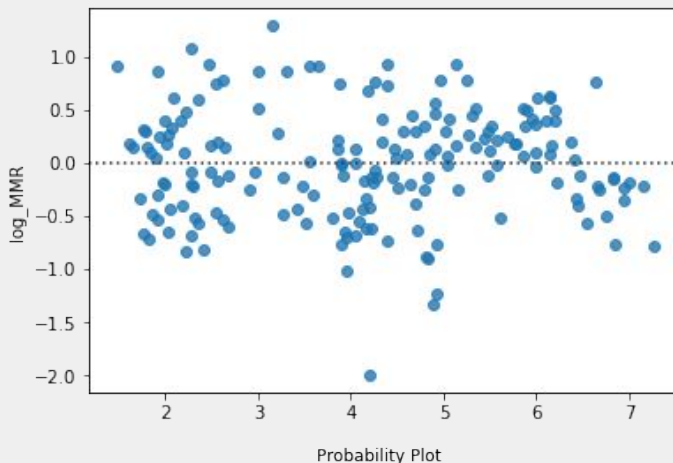
Interactions due to Continent seem to be non-linear for each feature, but most prominently with Internet Use



Interpreting Results

Dep. Variable:	log_MMR	R-squared:	0.902
Model:	OLS	Adj. R-squared:	0.897
Method:	Least Squares	F-statistic:	177.3
Date:	Sun, 21 Apr 2019	Prob (F-statistic):	1.82e-82
Time:	22:10:29	Log-Likelihood:	-138.57
No. Observations:	183	AIC:	297.1
Df Residuals:	173	BIC:	329.2
Df Model:	9		
Covariance Type:	nonrobust		

	Model	e	Percentage
Teen_fertility	0.004472	1.004482	0.448
Immunization_measles	-0.017785	0.982372	-1.763
Internet_use	-0.016604	0.983534	-1.647
log_GDP	-0.397712	NaN	-0.398
North_America	-0.431834	0.649317	-35.068
Oceania	-0.728902	0.482438	-51.756
Asia	-0.686526	0.503322	-49.668
Europe	-2.430763	0.087970	-91.203
IUxEurope	0.019884	1.020083	2.008



The Model has a decent R-squared with residuals that are normal and homoscedastic.

Interpretations of coefficients:

- For every 1 unit increase in Teen Pregnancy, Maternal Mortality Ratio increases .4%
- 1 unit increase Immunization decreases MMR 2%
- 1 unit increase Internet Use decreases MMR 2%
- For 1% increase in GDP per capita, MMR decreases .4%
- A country in Asia (vs South America or Africa) decreases MMR 50%
- Europe decreases MMR 91%
- North America decreases MMR 35%
- Oceania decreases MMR 52%