The effect of GDP on road traffic mortality

INTRODUCTION

Looking beyond the generalization that residents of high-income countries enjoy a better quality of life than people in low-income areas, this paper examines the relationship of GDP per capita with road traffic mortality, a leading cause of preventable death globally. Based on the WHO's 2015 Global status report, road traffic injuries were the number one cause of death globally in 2012 among those aged 15-29 (WHO, 2015). As well as being a public health problem, road traffic injuries are a development issue: low- and mid-income countries lose approximately 3% of GDP as a result of road traffic crashes.

a/ Based on Yannis et al. (2014), road safety is a concept correlated with mobility development in a country, which is in turn affected by socioeconomic factors such as the level of motorization, economic growth, etc.
b/ Most studies on the topic combine indicators of annual changes of GDP per capita level, annual changes of traffic mortality rate and variables signaling the general state of a country's transportation and healthcare infrastructure in a given year.
c/ Bishai et al. (2006) states that traffic fatalities increase with GDP per capita in lower income countries and decrease with GDP per capita in higher income countries.

Building on these results, my hypotheses were the followings:

H1: There was a significant relationship between road traffic mortality and GDP per capita in 2013.
H2: Regressing traffic mortality on GDP per capita, urban growth variable qualifies as a confounder.
H3: Factors influence traffic mortality rate differently in developed countries than in emerging ones.

DATA

The data was collected from the World Bank’s World Development Indicators website, downloaded on 19th December 2016, 16:19. After cleaning the original dataset of 217 observations and 6 variables, the final dataset had 173 observations and 12 variables. **The original variables were country name, traffic mortality rate, GDP per capita and selected potential confounders signaling the general state of a country's transportation and healthcare infrastructure in a given year (life expectancy, urban growth rate, continents).** Observations with missing values of the traffic mortality rate (37 observations) and of GDP per capita (27 observations) variables were dropped, the variable on GDP per capita was divided by 1,000 and was transformed into a new variable by taking its natural logarithm, and the categorical continent variable was transformed into binary variables for each continent. Description of the data is to be found in Appendix I, while descriptive statistics of the used variables are to be found in Appendix II. In Appendix III regression of traffic mortality rate on all explanatory variables are described and visualized, so that we can understand traffic mortality differences across different groups of countries. Main basic findings were the followings:

- Average traffic mortality rate was 26.17 people amongst 100,000 in Africa, 7.47 in Europe, 8.25 in North-America, 15.93 in South-America, 9.04 in Oceania and 16.24 in Asia in 2013.
- Annual urban growth has a positive, life expectancy has a negative and ln GDP per capita also has a negative average relationship with traffic mortality rate with a 5% statistical significance.
- Countries with an annual urban growth of 0.0% face a traffic mortality rate of 9.9 people per a population of 100,000, on average. Countries with a 1.0% higher annual urban growth are expected to have 3.0 higher traffic mortality rate in a population of 100,000, on average.

- Countries with a 1 year higher life expectancy are expected to have 0.76 lower traffic mortality rate in a population of 100,000, on average.
- Countries with an ln GDP per capita of 0.0% (such as Mozambique, Malawi or Niger) face a traffic mortality rate of 27.61 people per a population of 100,000, on average. Countries with a 10.0% higher ln GDP per capita are expected to have 0.50 lower traffic mortality rate in a population of 100,000, on average.
- Based on the smallest AIC and the largest R^2 values, the better fitting functional forms of regressing the traffic mortality on the explanatory variables one by one are cubic form for urban population growth, quadratic form for life expectancy and linear form for ln GDP per capita, however the differences are marginal between the alternative functional forms for all.

METHODOLOGY

Throughout the entire analysis I used robust standard errors to handle heteroscedasticity.

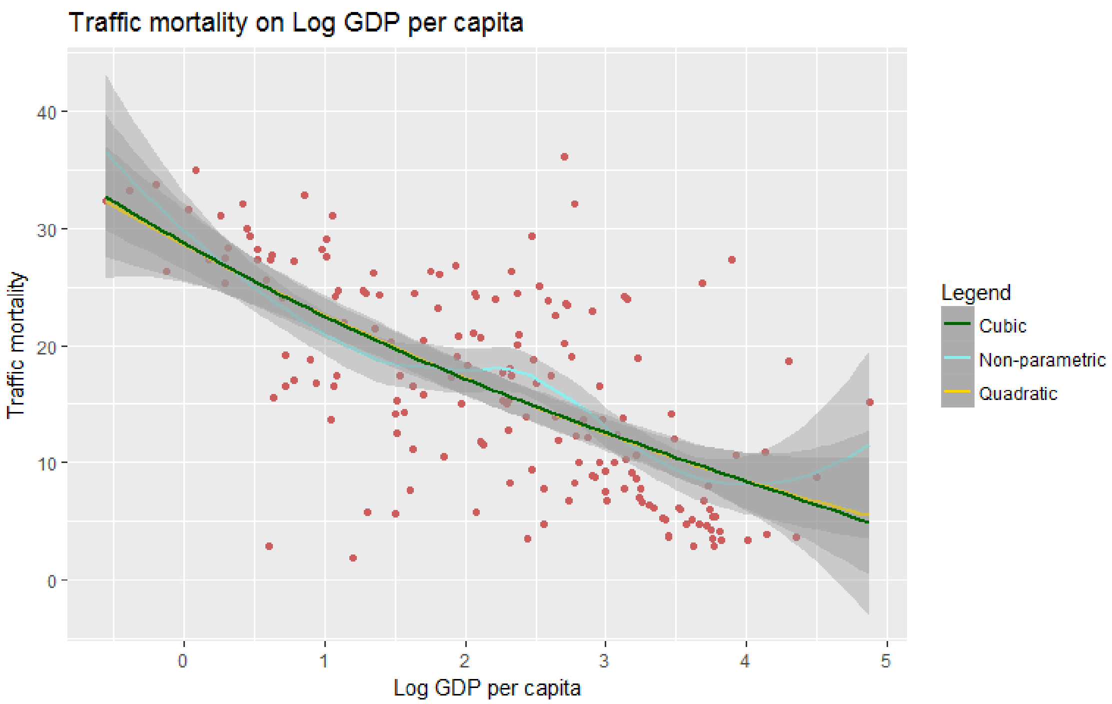
Deciding on **H1**, I examined the **relationship between traffic mortality and ln GDP per capita** at a 1.0% and 5.0% statistical significance by running an OLS regression and searching for an ideal functional form of ln GDP per capita.

Deciding on **H2**, I **examined possible confounders (life expectancy, urban growth, continent binaries)** in Appendix IV, which might take us closer to explain the differences in traffic mortality rate amongst countries. Since road safety differences can be composed of several factors, I chose to test possible confounders which are **surely correlated with ln GDP per capita, might be correlated with traffic mortality rate, but adding them to the regression model will not result in over controlling** for some of the differences across countries. I found life expectancy, urban growth and continent binaries good candidates, which could additionally explain differences in terms of general health of residents and quality of healthcare services (life expectancy), annual differences in population density and urban infrastructure (annual urban population growth), and socioeconomic differences between countries (continent binaries).

To be able to at least partially decide on **H3**, I added the **interaction terms of some selected confounding variables**. This way I could allow the slope of explanatory variables to differ based on the values of another explanatory variable, which helped to **further identify specific groups of countries with certain characteristics explaining their specific traffic mortality rate**. Even though I previously selected the better fitting functional form of the explanatory variables in the multiple regression, I finally decided to use them in the linear form. On the one hand, the functional forms of the explanatory variables did not provide a much better fit of the regression, therefore it was not worth it to add unnecessary complexity to the analysis. On the other hand, the interpretation of the interaction terms is significantly more straightforward when adding the explanatory variables in a linear form.

RESULTS

Regressing traffic mortality rate on ln GDP per capita, I found a relationship statistically significant at the 99% and 95% level. We can state with 95% confidence that countries with a 0.0% ln GDP per capita (such as Mozambique, Malawi or Niger) are expected to have a mortality rate of [25.23 ; 29.99] per a population of 100,000, in the general pattern that is represented by our data. We can state with 95% confidence that countries with a 10.0% higher ln GDP per capita are expected to have a mortality rate of [-0.58 ; -0.41] less, per a population of 100,000, in the general pattern that is represented by our data. This means that we can state with 95% confidence that countries with higher income are expected to have lower traffic mortality rates in the general pattern.



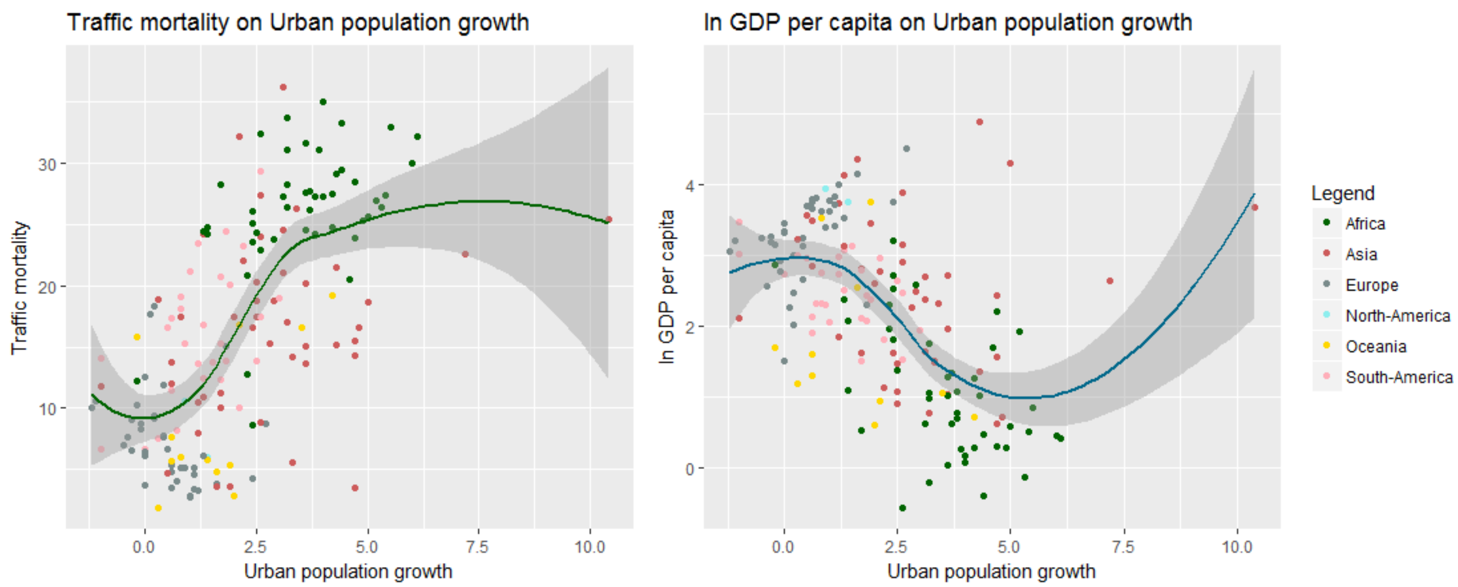
Nevertheless, we can see on the graph that there are several countries with large negative residuals around ln GDP per capita of 0.6, 1.5 and 2.5, while quite a few have large positive residuals at 2.5 and above 3.5. **Countries with largest positive and negative residuals** are listed below. A positive residual means that the country has a higher actual traffic mortality rate than estimated in the general pattern that is represented by our data. These countries have higher rate of preventable death than their income would predict. A negative residual means the opposite: these countries have a lower traffic mortality than their GDP per capita would predict in our model.

Country	Continent	Neg Resid	Traffic mortality rate
Kiribati	Oceania	-21.70	2.9
Micronesia, Fed. Sts.	Oceania	-19.76	1.9
Marshall Islands	Oceania	-15.39	5.7
West Bank and Gaza	Asia	-14.52	5.6
Tonga	Oceania	-12.03	7.6
Maldives	Asia	-11.97	3.5
Fiji	Oceania	-11.47	5.8
Palau	Oceania	-10.10	4.8
Afghanistan	Asia	-8.97	15.5
Bangladesh	Asia	-8.80	13.6

Country	Continent	Pos Resid	Traffic mortality rate
Thailand	Asia	22.04	36.2
Saudi Arabia	Asia	19.15	27.4
Iran, Islamic Rep.	Asia	18.29	32.1
Oman	Asia	16.13	25.4
Dominican Republic	South-Am.	14.01	29.3
Kuwait	Asia	12.52	18.7
Kazakhstan	Asia	12.19	24.2
Malaysia	Asia	12.08	24.0
Qatar	Asia	11.86	15.2
Jordan	Asia	10.25	26.3

Most countries with top negative residuals are from Oceania, where motorization is relatively less than ln GDP per capita would predict. Most countries with top positive residuals are from specific regions of Asia such as the Middle-East and South-East Asia.

Deciding on which control variables to include in the multiple regression, I examined life expectancy, urban growth and continent binaries as candidates by estimating their relationship with both traffic mortality rate and ln GDP per capita. With interpretations of the slope coefficients to be found in Appendix IV, by regressing ln GDP per capita on all potential confounder variables we can draw the conclusion that the **magnitude of the coefficient in case of urban growth, the africa binary and the europe binary variables is significantly larger than in the case of life expectancy at a 95% confidence level**. Keeping the regression of traffic mortality rate on ln GDP per capita as simple as possible, I decided that the urban growth and the continent variables should be the ones included as control variables and possible confounders. As described in the introduction, previous research on road safety often included variables describing the dynamics of motorization and transportation infrastructure development. I had the expectation towards urban growth to fulfill a similar function in this case. Additionally, by adding urban growth to the regression I was able to compare countries with the same urban population growth characteristics, but different in their country income. The graphs below showcase that **while countries form relatively distinct, visible clusters when regressing traffic mortality on urban population growth, they are more randomly distributed in the regression of ln GDP per capita on urban population growth**. Since countries with the same ln GDP per capita largely differ in urban population growth, this brings the **need to control for urban population growth also**, when adding ln GDP per capita as a control variable.



Searching for the regression model explaining most of the effect of country income on traffic mortality rate, I **experimented with adding the interaction terms in multiple ways**. This way I could allow the slope of urban population growth to differ by the continent binary variables or by ln GDP per capita. In regression (4) and (5) I interacted urban population growth with the europe and asia binary variables. In regression (6) I interacted urban population growth with ln GDP per capita. I had the africa binary variable as the reference category for all regressions which included continent binaries. The results of the regressions are showcased in Appendix V. By adding the control variables besides ln GDP per capita the slope coefficient of ln GDP per capita became smaller in absolute value and remained statistically significant at the 99% and 95% level in all regressions.

The coefficient of ln GDP per capita is the smallest in regression (4), where urban growth rate is interacted with the europe continent binary. We can state with 95% confidence the followings, in the general pattern that is represented by our data:

- Comparing countries from the same continent with the same urban population growth, the country with a 10.0% larger ln GDP per capita is expected to have [-0.11 ; -0.30] lower traffic mortality rate per 100,000 people.
- Comparing countries with the same ln GDP per capita, being outside of Europe, the country with a 1.0% higher urban population growth is expected to have [0.60 ; 1.84] lower traffic mortality rate per 100,000 people.
- Countries in Africa with 0.0% ln GDP per capita and 0.0% urban population growth are expected to have [21.19 ; 27.31] traffic mortality rate per 100,000 people.
- Countries in Asia with the same ln GDP per capita and the same urban population growth are expected to have [-3.54 ; -8.38] less traffic mortality rate per 100,000 people than countries in Africa with the same attributes. Similar interpr. of South-America, Oceania and North-America.
- Countries in Europe with the same ln GDP per capita and 0.0% urban population growth are expected to have [-6.16; -13.04] less traffic mortality rate per 100,000 people than countries in Africa with the same attributes.

In the data, the traffic mortality rate of European countries with the same ln GDP per capita, which have a 1.0% higher urban population growth is -0.64 lower, on average, than the traffic mortality rate of European countries with a 1.0% lower urban population growth. For non-European countries with the same ln GDP per capita, the corresponding difference of traffic mortality rate is 1.22, due to a 1.0% lower urban population growth. Comparing European countries with the rest of the world, with the same ln GDP per capita, European countries with a 1.0% urban population growth difference are expected to have -1.86 lower traffic mortality rate difference per 100,000 people, on average, than the traffic mortality rate difference in non-European countries. **This interpretation of the interaction term means that the effect of urban population growth on average traffic mortality rate is different across European and non-European countries, therefore we can state with a 90% confidence that countries inside and outside of Europe with the same level of urban population growth face different effects on traffic mortality rate.** The traffic mortality rate in Europe of countries with the same income is expected to be lower due to a higher urban growth rate, while the traffic mortality rate outside of Europe with the same income is expected to be higher due to a higher urban growth rate. **This eventually means that the growth of urban population leads to safer road transportation in Europe, while a less safe road transportation outside of Europe, based on the interaction term having a 10% statistical significance.**

By adding the interaction terms, we can conclude that the only statistically significant difference in the slope of urban growth could be estimated based on the europe continent binary. In case of regression (5) the interaction term of urban growth and the asia binary was not significant, while in regression (6) the interaction term of ln GDP per capita and urban growth was significant, but the urban population growth variable itself was not significant anymore even at the 10% level. The R^2 is almost the same, 0.70 for regressions where the continent binaries were added, including regression (3) to (6).

CONCLUSION

As **H1** was quite straightforward, I accepted the first hypothesis, due to the relationship between traffic mortality rate and ln GDP per capita being the statistically significant at the 5% level. Even though **H2** was not a fully objective criteria, to the extent of this paper I accepted it as well, as the level of analysis follows the previous practice of the course. **H3** was a broad hypothesis, rather an exploratory one, which in my assessment cannot be fully accepted. However, the finding on the opposite effect of urban population growth on traffic mortality in Europe and the rest of the world would worth further analysis, especially because it reflects similar results as other research projects on the topic.

REFERENCES

Yannis, G., Papadimitriou, E. and Folla, K. (2014). Effect of GDP changes on road traffic fatalities. *Safety Science*, 63:42-49.

World health Organization (2015). *Global status report on road safety 2015*, pp. 10.11.

Bishai, D., Quresh, A., James, P. and Ghaffar, A. (2006). National road casualties and economic development. *Health Economics*, 15(1):65-81.

APPENDIX I| Description of the dataset

Source	The data was collected from the World Bank's World Development Indicators website, downloaded on 19th December 2016, 16:19. This data can be considered as original of these countries, not experimentally controlled, containing information on actual mortality rate in a real-life environment.
Structure	The data was downloaded in a .csv format for each variable and then merged into a unified dataset (called *traffic_mortality.csv*). A continent column was added to the original data from the World Development Indicators website (new dataset called *traffic_mortality_mod.csv*). The dataset at this point was cross-sectional, where each row represented a separate country and columns represented variables describing features of the countries is 2013.
Data cleaning	The original dataset called tr consisted of 217 observations and 6 variables. Observations with missing values of the traffic mortality or GDP per capita variables were dropped. The variable on GDP per capita was divided by 1,000 and was transformed into a new variable by taking its natural logarithm. After cleaning the data and transforming variables the new subset of tr had 173 observations and 6 variables. Finally, the categorical continent variable was transformed into binary variables for each continent. After this transformation, the tr dataset had 173 observations and 12 variables.
Variables	***country*** • Name of the country • Format: String • No changes to the variable ***traf*** • Mortality caused by road traffic injury is estimated road traffic fatal injury deaths per 100,000 population. • Format: Numeric with one decimal • No changes to the variable ***urbgr*** • Urban population growth (annual %), referring to people living in urban areas as defined by national statistical offices. It is calculated as a weighted average, using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects. • Format: Numeric with one decimal • No changes to the variable ***life expectancy*** • Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life, calculated as a weighted average. • Format: Numeric with one decimal

- No changes to the variable

continent

- Continent the country is located in
- Format: Categorical
- Added to the original WDI dataset

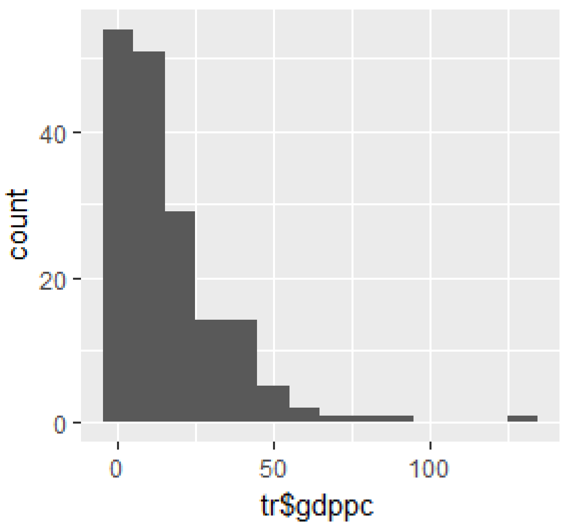
ln gdppc

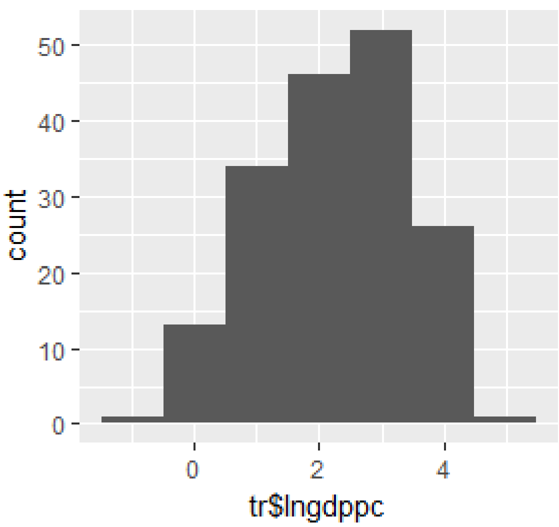
- Log GDP per capita, PPP (constant 2011 international $)
- Format: Natural logarithm, Numeric
- Created by taking the natural logarithm of GDP per capita ***(gdppc)***

africa, europe, northam, southam, oceania, asia

- Whether the country is located in the given continent
- Format: Binary (YES = 1 and NO = 0)
- Created from categorical ***continent*** variable

The original GDP per capita variable was substituted with its natural logarithm for two reasons. Firstly, GDP per capita differences between countries make more sense in relative than in absolute terms. Secondly, the variable had a positive skewness (2.40) with a long right tail, therefore its natural logarithm made its distribution to be closer to normal. The distributions of the original and the transformed variable can also be seen on the graphs below.

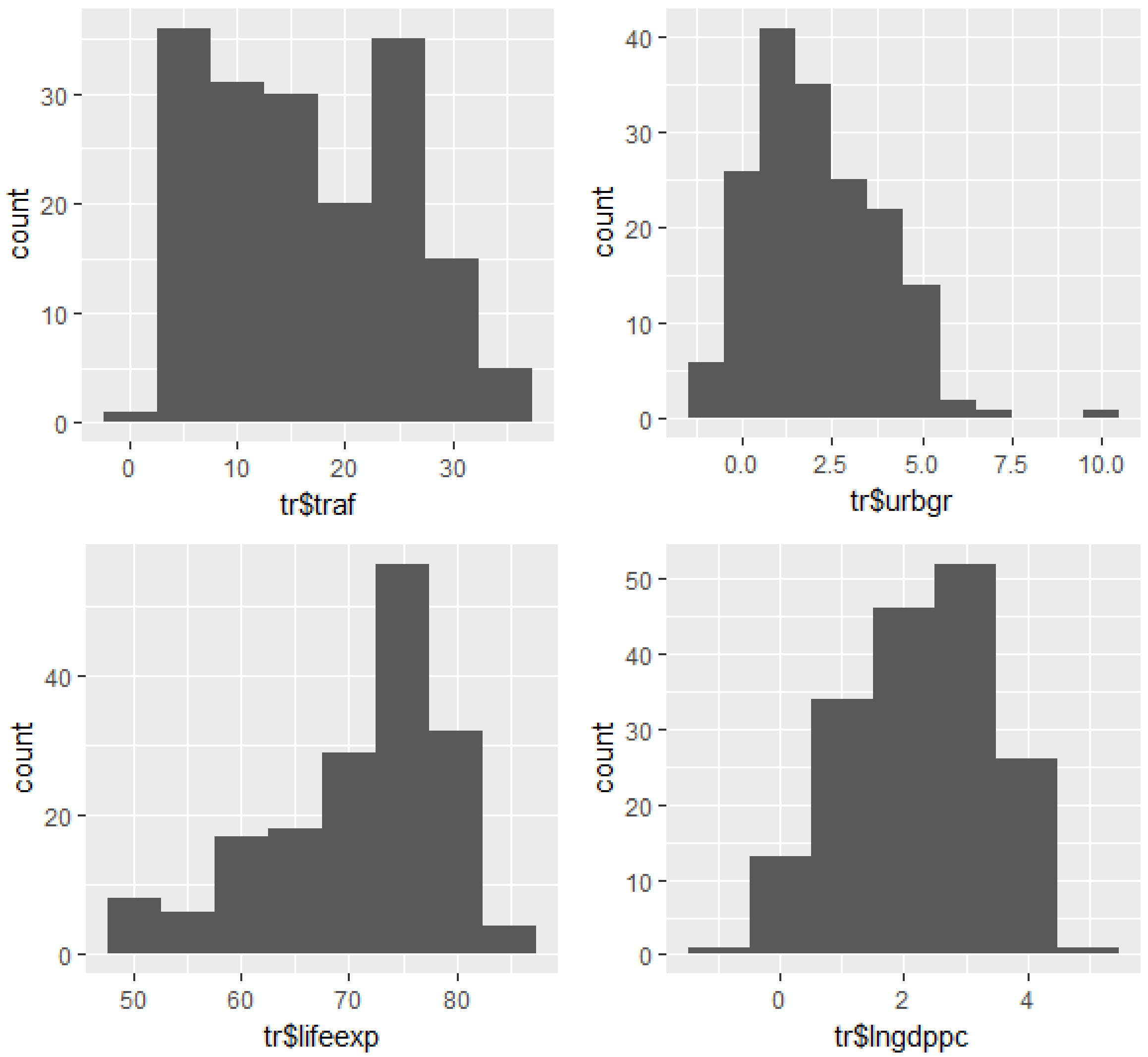




APPENDIX II | Descriptive Statistics

Table 1: Descriptive statistics of variables used

Variables	N	Min	Max	Mean	Median	Std Dev	Skewness
traf	173	1.90	36.2	16.46	16.60	8.94	0.17
urbgr	173	-1.20	10.40	2.14	1.90	1.80	0.80
lifeexp	173	48.90	83.30	70.95	73.20	8.52	-0.70
ln gdppc	173	-0.56	4.88	2.24	2.37	1.17	-0.23



Even though traffic mortality is slightly positively skewed, and its meaning would make more sense in relative than in absolute terms (percentage difference of traffic mortality rate instead of number of people per 100,000 individuals), I decided not to take the natural logarithm of the variable, since it does not have a long right tail, its distribution is already close enough to normal.

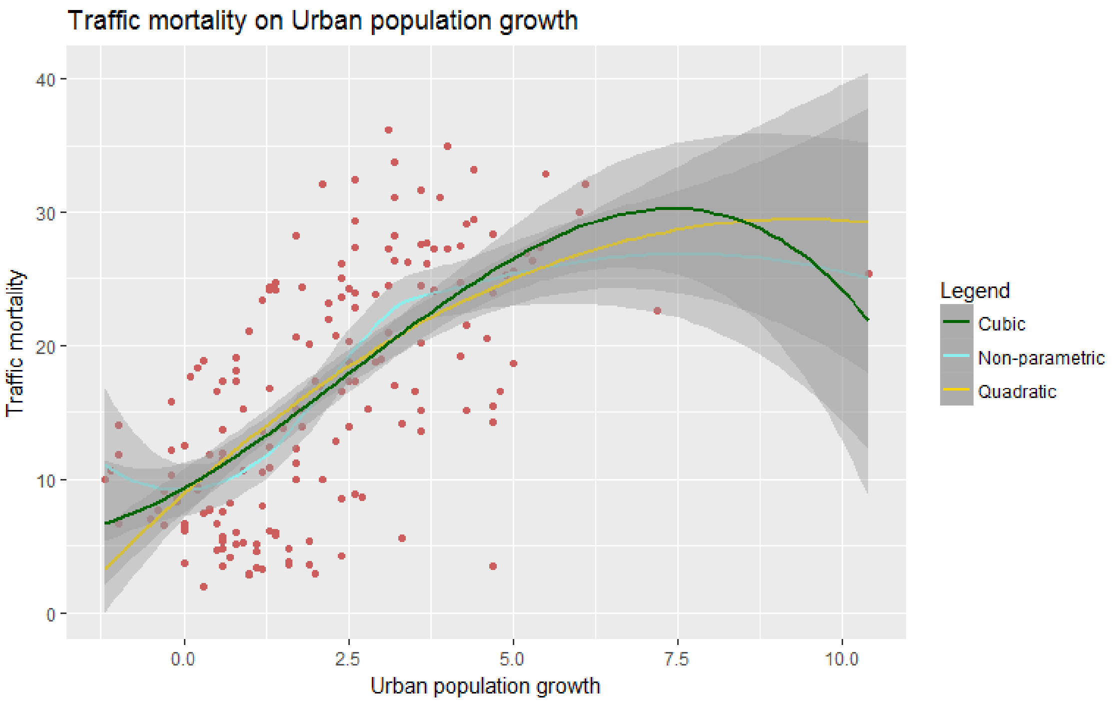
APPENDIX III | Traffic mortality differences

Traffic mortality differences across continents

To calculate the differences in traffic mortality across continents, new datasets were defined for observations belonging to each continent. For each dataset, means of the traffic mortality variable were calculated by regressing traffic mortality on the continent binaries, where the intercept can be interpreted as the expected traffic mortality on average in the dataset of all continents except for the one in the regression, and the slope coefficient can be interpreted as the expected average difference in traffic mortality between the continent in the regression and the rest of the world.

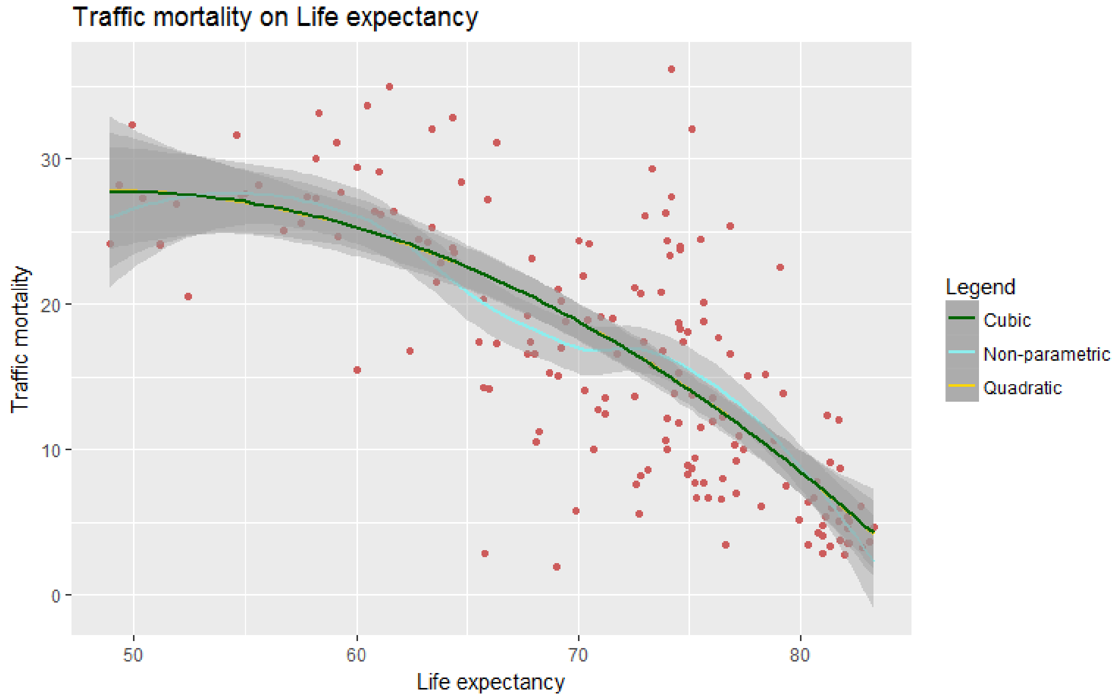
Continent	Mean	p25	p50	p75	p95
Africa	26.17	24.25	26.40	28.75	33.10
Europe	7.47	4.65	6.65	9.20	15.50
North-America	8.25	7.15	8.30	9.45	10.40
South-America	15.93	12.40	15.30	19.10	24.00
Oceania	9.04	5.25	5.90	16.00	17.90
Asia	16.24	11.80	16.60	21.00	27.20

Traffic mortality differences by urban population growth



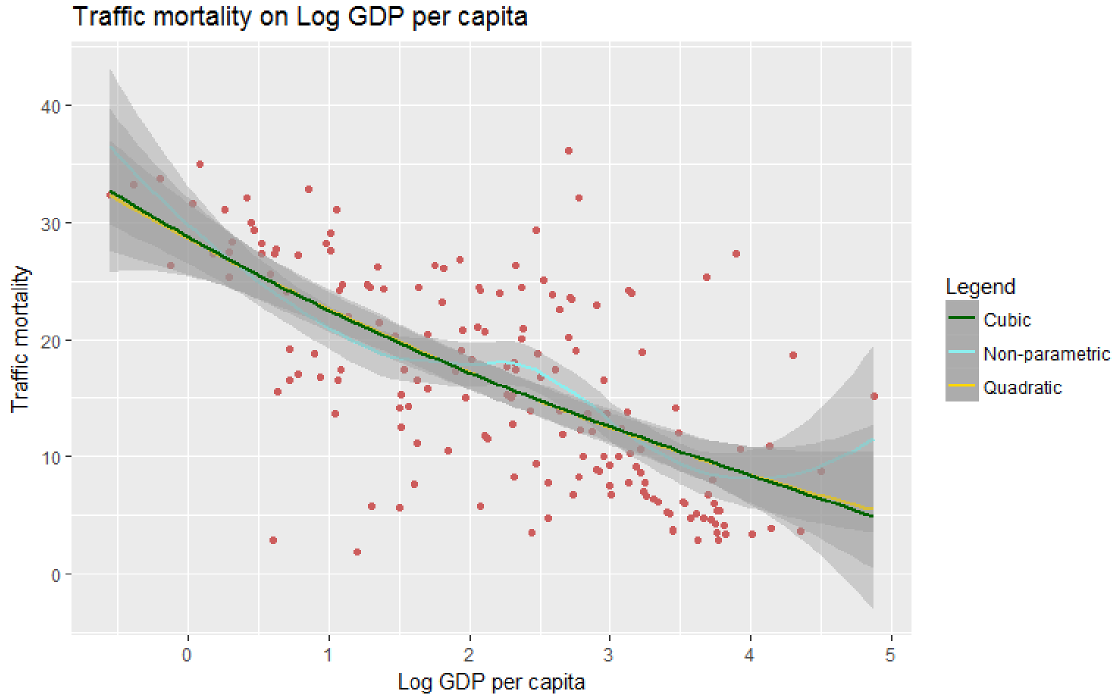
Traffic mortality on Urban population growth	Linear	Quadratic	Cubic
R^2	0.38	0.40	***0.41***
AIC	1151.36	1147.22	***1145.46***

Traffic mortality differences by life expectancy



Traffic mortality on Life expectancy	Linear	Quadratic	Cubic
R^2	0.52	***0.56***	0.56
AIC	1105.75	***1095.97***	1097.96

Traffic mortality differences by ln GDP per capita



Traffic mortality on ln GDP per capita	Linear	Quadratic	Cubic
R^2	***0.45***	0.43	0.43
AIC	***1158.29***	1159.34	1161.30

APPENDIX IV | Regressing possible confounders on ln GDP per capita

	Dependent variable:							
	lngdppc							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
urbgr	-0.32^{***}							
	(0.04)							
lifeexp		0.11^{***}						
		(0.01)						
africa			-1.52^{***}					
			(0.16)					
europe				1.31^{***}				
				(0.19)				
northam					1.62^{*}			
					(0.83)			
southam						0.29		
						(0.24)		
oceania							-0.52	
							(0.35)	
asia								0.27
								(0.20)
Constant	2.92^{***}	-5.66^{***}	2.65^{***}	1.95^{***}	2.22^{***}	2.19^{***}	2.28^{***}	2.17^{***}
	(0.12)	(0.45)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)
N	173	170	173	173	173	173	173	173
R^2	0.24	0.65	0.34	0.22	0.02	0.01	0.01	0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Based on regressions (1) to (8) we can conclude that urban growth, life expectancy, the africa binary and the Europe binary have a statistically significant relationship with ln GDP per capita at a 1.0% and 5.0% level, on average. The rest of the binary continent variables do not have a statistically significant relationship with ln GDP per capita at a 1.0% or 5.0% level, on average.

We can state with 95% confidence that countries with a 0.0% annual urban growth are expected to have a [2.68 ; 3.16] percent ln GDP per capita, in the general pattern that is represented by our data. We can state with 95% confidence that countries with a 1.0% higher annual urban growth are expected to have a [-0.40 ; -0.24] percentage point lower ln GDP per capita, in the general pattern that is represented by our data.

We can state with 95% confidence that countries with a 1 year higher life expectancy are expected to have a [0.09 ; 0.13] percentage point higher ln GDP per capita, in the general pattern that is represented by our data.

We can state with 95% confidence that countries in Africa are expected to have a [-1.84 ; -1.20] percentage point lower ln GDP per capita than the rest of the world, in the general pattern that is represented by our data.

We can state with 95% confidence that countries in Europe are expected to have a [0.93 ; 1.69] percentage point higher ln GDP per capita than the rest of the world, in the general pattern that is represented by our data.

It is visible that the magnitude of the coefficient in case of the urban growth, the africa binary and the europe binary variables is significantly larger than in the case of life expectancy at a 95% confidence level.

APPENDIX V | Multiple Regressions with control variables and interactions

	Dependent variable:					
	traf					
	(1)	(2)	(3)	(4)	(5)	(6)
lngdppc	-4.98***	-3.52***	-2.29***	-2.10***	-2.28***	-3.66***
	(0.44)	(0.46)	(0.46)	(0.47)	(0.47)	(0.82)
urbgr		1.94***	1.04***	1.22***	1.08**	-0.06
		(0.30)	(0.29)	(0.31)	(0.43)	(0.62)
asia			-5.84***	-5.96***	-5.58**	-6.36***
			(1.22)	(1.21)	(2.16)	(1.23)
europe			-10.65***	-9.60***	-10.53***	-10.42***
			(1.62)	(1.72)	(1.82)	(1.61)
northam			-9.19**	-9.28**	-9.10**	-8.59**
			(3.89)	(3.87)	(3.94)	(3.86)
oceania			-13.66***	-13.42***	-13.57***	-14.55***
			(1.75)	(1.74)	(1.84)	(1.79)
southam			-4.78***	-4.63***	-4.69***	-5.25***
			(1.45)	(1.45)	(1.58)	(1.46)
urbgr:europe				-1.86*		
				(1.07)		
urbgr:asia					-0.08	
					(0.58)	
lngdppc:urbgr						0.45**
						(0.22)
Constant	27.61***	20.20***	25.09***	24.25***	24.92***	29.04***
	(1.12)	(1.51)	(1.46)	(1.53)	(1.87)	(2.43)
Observations	173	173	173	173	173	173
R^2	0.43	0.54	0.69	0.70	0.69	0.70

Note: *p<0.1; **p<0.05; ***p<0.01

In regression (6) ln GDP per capita is interacted with urban growth rate. We can state with 95% confidence that when comparing countries from the same continent with 0.0% urban population growth, the country with a 10.0% larger ln GDP per capita is expected to have [-0.53 ; -0.20] lower traffic mortality rate per 100,000 people, in the general pattern that is represented in our data. However, in regression (6) urban growth rate is not statistically significant even at the 10% level, probably due to the interaction term being significant at the 95% level, therefore regression (6) does not explain more of the effect we are after. In regression (5) the interaction term is not significant.

APPENDIX VI | Code

```
install.packages ("pastecs")
install.packages("DataCombine")
install.packages("descr")
install.packages("varhandle")
install.packages("knitr")
install.packages("gmodels")

library(knitr)
library(gmodels)
library(pastecs)
library(DataCombine)
library(descr)
library(arm)
library(readr)
library(dplyr)
library(ggplot2)
library(lmtest)
library(sandwich)
library(segmented)
library(splines)
library(stargazer)
library(fBasics)
library(varhandle)

# CLEAR MEMORY
rm(list=ls())

# SET WORKING DIRECTORY
setwd("C:/Users/Keresztesi Luca/Box Sync/CEU 1st trimester/Data Analysis 2/Term Project")
getwd()

# LOAD  DATA
tr<- read.csv2("traffic_mortality_mod.csv",
              header = TRUE,
              dec = ".",
              strip.white = TRUE,
              na.strings = "..",
              stringsAsFactors = FALSE)

# CLEAN AND CREATE VARIABLES
tr <- subset(tr, tr$traf !="NA" & tr$gdppc !="NA")
tr$gdppc <- tr$gdppc/1000
tr$lngdppc <- log(tr$gdppc)
tr$gdppc <- NULL

tr$africa <- tr$continent == "Africa"
tr$europe <- tr$continent == "Europe"
tr$northam <- tr$continent == "North-America"
tr$southam <- tr$continent == "South-America"
tr$oceania <- tr$continent == "Oceania"
tr$asia <- tr$continent == "Asia"

# DESCRIPTIVE STATS
basicStats(tr[,2])
qplot(tr$traf, geom="histogram", binwidth=5)
basicStats(tr[,3])
qplot(tr$urbgr, geom="histogram", binwidth=1)
basicStats(tr[,4])
qplot(tr$lifeexp, geom="histogram", binwidth=5)
basicStats(tr[,6])
qplot(tr$lngdppc, geom="histogram", binwidth=1)

# LOWESS NONPARAMETRIC REGRESSION
ggplot(data = tr, aes(x=urbgr, y=traf)) +
  ggtitle("Traffic mortality on Urban population growth") + xlab("Urban population growth") + ylab("Traffic 
mortality") +
  geom_point(size=1.5, aes(colour=factor(continent)))+
  geom_smooth(method="loess", colour="darkgreen")+
  scale_colour_manual(name="Legend",values=c("darkgreen","indian red","lightcyan 4","darkslategray 2","gold 
1","lightpink 1"))

ggplot(data = tr, aes(x=urbgr, y=lngdppc)) +
  ggtitle("ln GDP per capita on Urban population growth") + xlab("Urban population growth") + ylab("ln GDP per 
capita") +
  geom_point(size=1.5, aes(colour=factor(continent)))+
  geom_smooth(method="loess", colour="deepskyblue4")+
  scale_colour_manual(name="Legend",values=c("darkgreen","indian red","lightcyan 4","darkslategray 2","gold 
1","lightpink 1"))
```

```
 # EXPLANATORY VARIABLES
 # URBAN GROWTH AND TRAFFIC MORTALITY
 reg_u <- lm(traf ~ urbgr, data=tr)
 summary(reg_u, vcov=sandwich)

 ggplot(data = tr, aes(x=urbgr, y=traf)) +

  ggtitle("Traffic mortality on Urban population growth") + xlab("Urban population growth") + ylab("Traffic mortality") +
  geom_point(size=1.5, colour="indian red")+
  geom_smooth(method="loess", aes(x=urbgr, y=traf, colour="Non-parametric")) +
  geom_smooth(method="lm", formula=y~poly(x,2), aes(x=urbgr, y=traf, colour="Quadratic")) +
  geom_smooth(method="lm", formula=y~poly(x,3), aes(x=urbgr, y=traf, colour="Cubic")) +
  scale_colour_manual(name="Legend",values=c("darkgreen", "darkslategray 2","gold 1"))

reg_u_2 <- lm(traf ~ poly(urbgr,2), data=tr)
summary(reg_u_2, vcov=sandwich)
reg_u_3 <- lm(traf ~ poly(urbgr,3), data=tr)
summary(reg_u_3, vcov=sandwich)
AIC(reg_u, reg_u_2, reg_u_3)

# LIFE EXPECTANCY AND TRAFFIC MORTALITY
reg_l <- lm(traf ~ lifeexp, data=tr)
summary(reg_l, vcov=sandwich)

ggplot(data = tr, aes(x=lifeexp, y=traf)) +
  ggtitle("Traffic mortality on Life expectancy") + xlab("Life expectancy") + ylab("Traffic mortality") +
  geom_point(size=1.5, colour="indian red")+
  geom_smooth(method="loess", aes(x=lifeexp, y=traf, colour="Non-parametric")) +
  geom_smooth(method="lm", formula=y~poly(x,2), aes(x=lifeexp, y=traf, colour="Quadratic")) +
  geom_smooth(method="lm", formula=y~poly(x,3), aes(x=lifeexp, y=traf, colour="Cubic")) +
  scale_colour_manual(name="Legend",values=c("darkgreen", "darkslategray 2","gold 1"))

reg_l_2 <- lm(traf ~ poly(lifeexp,2), data=tr)
summary(reg_l_2, vcov=sandwich)
reg_l_3 <- lm(traf ~ poly(lifeexp,3), data=tr)
summary(reg_l_3, vcov=sandwich)
AIC(reg_l, reg_l_2, reg_l_3)

# LN GDPPC AND TRAFFIC MORTALITY
reg_g <- lm(traf ~ lngdppc, data=tr)
summary(reg_g, vcov=sandwich)

ggplot(data = tr, aes(x=lngdppc, y=traf)) +
  ggtitle("Traffic mortality on Log GDP per capita") + xlab("Log GDP per capita") + ylab("Traffic mortality") +
  geom_point(size=1.5, colour="indian red")+
  geom_smooth(method="loess", aes(x=lngdppc, y=traf, colour="Non-parametric")) +
  geom_smooth(method="lm", formula=y~poly(x,2), aes(x=lngdppc, y=traf, colour="Quadratic")) +
  geom_smooth(method="lm", formula=y~poly(x,3), aes(x=lngdppc, y=traf, colour="Cubic")) +
  scale_colour_manual(name="Legend",values=c("darkgreen", "darkslategray 2","gold 1"))

reg_g_2 <- lm(traf ~ poly(lngdppc,2), data=tr)
summary(reg_g_2, vcov=sandwich)
reg_g_3 <- lm(traf ~ poly(lngdppc,3), data=tr)
summary(reg_g_3, vcov=sandwich)
AIC(reg_g, reg_g_2, reg_g_3)

# CONTINENT AND TRAFFIC MORTALITY
traf_af <- subset(tr, tr$continent == "Africa")
quantile(traf_af$tr, c(.25, .50,  .75, .95))
traf_eu <- subset(tr, tr$continent == "Europe")
quantile(traf_eu$tr, c(.25, .50,  .75, .95))
traf_na <- subset(tr, tr$continent == "North-America")
quantile(traf_na$tr, c(.25, .50,  .75, .95))
traf_sa <- subset(tr, tr$continent == "South-America")
quantile(traf_sa$tr, c(.25, .50,  .75, .95))
traf_oc <- subset(tr, tr$continent == "Oceania")
quantile(traf_oc$tr, c(.25, .50,  .75, .95))
traf_as <- subset(tr, tr$continent == "Asia")
quantile(traf_as$tr, c(.25, .50,  .75, .95))
```

```
reg_af <- lm( traf ~ africa, data=tr)
summary(reg_af, vcov=sandwich)
reg_eu <- lm(traf ~ europe, data=tr)
summary(reg_eu, vcov=sandwich)
reg_na <- lm(traf ~ northam, data=tr)
summary(reg_na, vcov=sandwich)
reg_sa <- lm(traf ~ southam, data=tr)
summary(reg_sa, vcov=sandwich)
reg_oc <- lm(traf ~ oceania, data=tr)
summary(reg_oc, vcov=sandwich)
reg_as <- lm(traf ~ asia, data=tr)
summary(reg_as, vcov=sandwich)

# LINEAR REG OF LNGDPPC ON URBAN GROWTH AND CONTINENTS
reg1_urb <- lm(lngdppc ~ urbgr, data=tr)
summary(reg1_urb, vcov=sandwich)
reg1_lif <- lm(lngdppc ~ lifeexp, data=tr)
summary(reg1_lif, vcov=sandwich)
reg1_af <- lm(lngdppc ~ africa, data=tr)
summary(reg1_af, vcov=sandwich)
reg1_eu <- lm(lngdppc ~ europe, data=tr)


summary(reg1_eu, vcov=sandwich)
reg1_na <- lm(lngdppc ~ northam, data=tr)
summary(reg1_na, vcov=sandwich)
reg1_sa <- lm(lngdppc ~ southam, data=tr)
summary(reg1_sa, vcov=sandwich)
reg1_oc <- lm(lngdppc ~ oceania, data=tr)
summary(reg1_oc, vcov=sandwich)
reg1_as <- lm(lngdppc ~ asia, data=tr)
summary(reg1_as, vcov=sandwich)

stargazer(list(reg1_urb, reg1_lif, reg1_af, reg1_eu, reg1_na, reg1_sa, reg1_oc, reg1_as), digits=2,
out="traffic_mortality_1.html")

# RESIDUALS
reg_g_predict <- predict(reg_g)
tr$e_lngdppc <- resid(reg_g)
basicStats(tr[,13])
qplot(tr$e_lngdppc, geom="histogram", binwidth=3)
# countries with most negative residuals
e_neg_lngdppc <- subset(tr, e_lngdppc<(-8.0))
# countries with most positive residuals
e_pos_lngdppc <- subset(tr, e_lngdppc>(10.0))

# LINEAR REG OF OF TRAFFIC MORTALITY ON LNGDPPC
reg1 <- lm(traf ~ lngdppc, data=tr)
summary(reg1, vcov=sandwich)

reg2 <- lm(traf ~ lngdppc + urbgr, data=tr)
summary(reg2, vcov=sandwich)

reg3 <- lm(traf ~ lngdppc + urbgr + asia + europe + northam + oceania + southam, data=tr)
summary(reg3, vcov=sandwich)

reg4_1 <- lm(traf ~ lngdppc + urbgr + asia + europe + northam + oceania + southam + europe*urbgr, data=tr)
summary(reg4_1, vcov=sandwich)

reg4_2 <- lm(traf ~ lngdppc + urbgr + asia + europe + northam + oceania + southam + asia*urbgr, data=tr)
summary(reg4_2, vcov=sandwich)

reg5 <- lm(traf ~ lngdppc + urbgr + asia + europe + northam + oceania + southam + urbgr*lngdppc, data=tr)
summary(reg5, vcov=sandwich)

stargazer(list(reg1, reg2, reg3, reg4_1, reg4_2, reg5), digits=2, out="traffic_mortality_2.html")
```