Explaining suicide in Italian counties

Pietro Romozzi & Federico Colombo

20 novembre 2014

Abstract:

With the present work, the authors intend to create a descriptive model of economical, social and climatic characteristics of a territory, aiming to assess the impact of these factors on suicide rates. The model will be tested on Italian counties (110 observations), couniting on data obtained by ISTAT.

Introduction:

Topic

Several papers have tried to propose an explanatory framework to describe how climate conditions affect individuals' mental health. Climate may affect it both directly through extreme climate change, which exposes people to trauma, or indirectly, as some climate's characteristics may affect physical activity and mental health. Extreme temperature or humidity rates have an influence on individuals' physical health as it becomes more energy consuming to do physical activity, to work or to travel, but also mental health could be negatively affected as climate erodes physical environments, which subsequently damage social environments and thus, mental condition. Health is defined by the World Health Organisation as a 'state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity' (World health organisation, 1948). Mental health is one of the key components that define general health and it can be seen as an alteration in thinking, mood, behaviour, and associated distress or impaired functioning. Usually, problems related to mental condition can be considered to be either uncommon (schizophrenia, anxiety and personal disorders) or common (depression, dementia, physiological distress). The WHO showed that among almost 90% of people who died by suicide in high-income countries at least one mental disorder was present, while those affected by more than one mental disorder were experiencing significantly higher risks of committing suicide. The relationship between mental health (and the related suicide rates) and climate still needs to be studied into details, but some scholars identified extreme heat as a factor that is often associated with a general increase in aggressive, criminal and suicide rates (Bouchama A. (2007)). Humidity has also been associated with mental functioning, especially with poorer concentration and elevated fatigue. Climate change is a phenomenon which is expected to have severe consequences on the mental health of millions of people too: the threat to physical health or the potential loss of home, jobs and familiars due to catastrophic climate changes negatively affects mental condition. Sunlight is another factor, which has been documented to have direct effects on climate related health problems such as depressive symptoms, anxiety disorders, and other mental problems usually peaking in the cold season when there is limited sunlight (Magnusson (2007)). The debate about the impact of climate change on human health has only very recently included considerations of mental health. Climate can have an influence on our psyche in a varieties of ways: it can directly inflict more and more natural disasters on human settlements, which, as evidence suggests, increases anxiety-related or chronic severe mental problems; climate can also increase the risk of physical injuries or health problems, which are correlated with mental health; finally, climate can endanger the natural and social environment on which people depend for their wellbeing.

Relevance

One obvious reason why this work is potentially interesting is that mental disorders and suicide are nowadays often discussed and debated both among insitution in health's field and media. Indeed, according to the latest WHO report concerning suicide prevention (see the latest report by WHO 2014), "Globally, among young adults 15-29 years of age suicide accounts for 8.5% of all deaths and is ranked as the second leading

cause of death (after traffic accidents). Among adults aged 30-49 years it accounts for 4.1% of all deaths and is ranked the fifth leading cause of death". This statement displays clearly the relevance of this issue, which is also greater considering that these estimation are often underrated because of the difficulties in recognizing sucide as a cause of death - will of committing suicide could be not recognized, families could prefer not to speak about it, and so on (see the note by ISTAT 2012). The above mentioned paper, moreover, is considered by several OECD countries as a strong stimulus to move further in suicide prevention. Among these countries, we are going to focus on Italy, particularly at provincial level. Our convintion is that a focus on smaller-than-countries environments, does help in assessing how factors affecting suicides impact differently within the same country, thus highlighting a possible variance among regions which is overlooked by country-level aggregate data. In addition to this "micro-level" focus, we also intend to stress some environmental and climatic characteristics which could contribute to create "suicide-genic" environment. In fact, according to a recently developed body of studies, there is a correlation between environmental-climatic characteristics and mental health. So, considering mental disorder as a possible driver to suicide, we include these types of variables in our study trying to widen the range of explanatory factors for the phenomenon. The above mentioned link is often not considered, as some auhors like Berry, Bowen, and Kjellstrom (2009) underline, therefore we believe that including them in the present paper could deliver interesting results and perhaps stimulate further research in this branch. We also do not neglet the use of more "ortodox" variables (i.e. related to economic and social dimensions) as control variables to develop a fairly comprehensive perspective. Looking for empirical evidence from Italian data, our aim is to add knowledge to existing literature reporting specific suicide patterns in Italy, among which we will look for evidence of the magnitude of climatic and environmental factors, believing that our wide-spectrum approach makes the present work interesting and innovative. Another interesting feature is the provincial focus which allows to underscore, in the end, contingent specific characteristics of the territory rather than the whole country. What is interesting is the the fact that provinces across Italy vary greatly with respect to both socioeconomic factors such as unemployment rate, income distribution, GDP per capita, the South being on average less wealthy and more unequal than the North, and also climatic characteristics differ from region to region. Recognizing these characteristics, will hopefully provide information to policy makers for setting efficient preventive strategies at a more manageable and flexible micro-level, serving as an input for tailored and focused polices.

Research questions:

As said above, this paper will focus on the relation between suicidal behaviours (an extreme consequence of mental disorders) and various factors concerning climate (Magnusson 2007) and socio-economic context.

Regarding *climate*, we want to test if the following hypoteses hold true:

- Suicide rate in a certain Italian county increases when the average temperature is low (Bouchama A. 2007).
- Suicide rate in a certain Italian county increases when the average temperature range is high
- Suicide rate increases when precipitations happen to be frequent.

Regarding the *socio-economical framework* instead, we want to verify the follwing hypoteses:

- Suicide rate increases when GDP per capita is high (OECD 2013).
- Suicide rate increases according to economic inequality (represented by the gini index).
- Suicide rate increases whereby unemployment rate is higher.
- Suicide rate is higher in more populated counties.

The hypotheses we want to test originate from a careful review of the relevant literature regarding mental disorders, suicidal patterns and the magnitude of climate impact on them.

Methodology:

Sources and data gathering

For the scope of this work, data for different variables were necessary for all of the Italian counties; in particular we looked for data about temperature, temperature range, precipitations, GDP, gini's index, unemployment, population, and naturally suicides, which is our depending variable. We decided to carry out our analysis on a single year, because of the higher complexity in data gathering process and given the time constraint this research is under. To have a significant result, we believed that recent data were necessary, so we identified 2009 as an acceptable compromise between proximity in time and availability of data for all of our variables, thus all data gathered and used regard 2009. Most of the data required were founded in the Italian National Institute of Statistics (ISTAT) database, and fortunately the missing ones were available in other databases of ISTAT-related agencies or in the Eurostat database. Accessing to the ISTAT database, which is similar to the OECD one, is free and easy, and data are clear and well structured. Nevertheless, we needed to process them in order to obtain so-called **tidy data**, required by R to analyze and modeling them. We recall a definition of tidy data, according to (Wickham 2014):

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table.

Given that, we firstly collected and downloaded all the data from ISTAT website (**http://www.istat.it/it/**) and from ISTAT-related agencies' databases. For practical reasons we carried out the tidying process in Microsoft Excel. Basically, we made sure that all names of different counties were the same in all datasets we downloaded (otherwise we simply renamed the counties), as well as checking that our values were expressed in the appropriate scale, further discussed below. For istance, we decided to rescale GDP per capita from Euro to thousands Euros, to facilitate the interpretation of our results. To sum up, we downloaded data and tidied them usinc Excel obtaining three dataframes, namely **gdp_data**, **socio_eco_data** and **climate_data**. Then we converted them in .csv format, easily readable by R, and finally we moved into RStudio to merge these three dataframes. In the end we collected 110 observations, one for each Italian county, of 9 variables.

Considered variables and analysis structure:

The variable *suicide rate* is defined by the WHO as the standardized number of deaths by suicide every 100.000 inhabitants occurred in a certain country. We collected data about suicide rate for every Italian county, from the ISTAT database.

We adopted the variable *population* as a proxy of the degree of urbanization of the area. It describes the total number of people living in the county. Carrying out our analysis, we decided to take value for population in logaritmic scale. Also in this case, we found the data about population in the ISTAT database.

Regarding the variable *GDP per capita*, we found in the Eurostat database a precious source. Nevertheless, the data provided were expressed in Euro, deemed not to provide significant impact in our linear analysis since a one-euro variation is certainly too small to impact on suicide rate. Therefore, using Excel during our data cleaning process, we expressed GDP per capita in thousands Euro, so that our coefficient will represent the impact on suicide rate caused by a variation of 1000 Euro, perhaps delivering more significant results.

The *Gini Coefficient*, is the most common measure of income inequality. It varies between 0 and 1 and expresses inequality among values of the frequency distribution of income for our target counties. Even these data were available via Eurostat, which provides the index either including households' income from the rent of owned properties or not. Since in Italy, has been culturally widespread to invest in real estate, at least until the economic crisis of 2008 and the subsequent introduction of a strong tax levied on real estate properties, we believe that adopting the Gini index including rent might better represent the italian context.

As expected, we also notice gini's index to be slightly higher in Southern counties compared to the Centre and the North.

We included in our analysis also *unemployment rate* since economical, social and psychological consequences of job loss might have an influence in mental disorders and eventually in suicidal behavior. This data, provided by ISTAT refers to the share of the labor force that is without work but available for and seeking employment.

The variable *average temperature* simply corresponds to the value of the annual average temperature in a certain county expressed in Celsius degrees. Insted, *average temperature range* corresponds to the average difference between the maximum and the minimum temperature registered in the county, again expressed in Celsius degrees. Then, *precipitations* measures the annual quantitiy of rain expressed in millimiters. Those data about climate are obtained from CRA-CMA, a ISTAT-related agency, particularly the department for meteorology applied to agriculture (ISTAT 2012).

Lastly, we would like to include also *solar radiation* measuring the level of solar irradiance in megajoul (MJ) per square meter. The Italian Air Force offers a metereological service, which is responsible for the measurement of solar irradiance on the Italian territory. Although we intended to use this variable, county-level data were not available, so eventually we dropped it.

We believe that a linear model was an effective way to analyze our dataframe because it could give a clear idea of the impact of the selected indipendent variables on suicides which we chose as to be the dependent variable. As the reader will observe below, we started from a linear model including all the variables we considered, then recognizing that some variables were particularly unsignificant, we opted to remove two of them obtaining a simplified model which appeared fairly significant and interesting. In the end we carried out our analysis providing two linear models, whose interpretation will be discussed further, then we enrich the statistical output with some visual displays, particularly focusing on the relationship between suicides and climatic factors.

All the data gathered will be processed using "R" (see R Core Team 2014). We use as well some packages developed for R, namely **ggplot2** (see Wickham and Chang 2014), **stargazer** (see **???**), **repmis** (see Gandrud 2014), **car** (see **???**) and **knitr** (see **???**).

Case studies:

In a previous work, we tried to test the model we will use in this analysis on regional data. However, we faced the big issue of not having enough observations, given that Italy counts 20 regions and it resulted in a statistically not significant model. Nevertheless, we were fairly confident in the goodness of our model and interested in giving an answer to our research question, we threfore decided to test it again on a sufficient number of observations. This led us to move our analysis to a county-level, thus having available 110 observations. Eventually, in the present work we take into consideration all the Italian counties.

Descriptive Statistics:

The table below presents the observations' ranging distributions of our dataset. As it can be seen, the number of observations ranges from 103 to 110 observations. This is due to the fact that exactly during the year taken into consideration, i.e. 2009, some new provinces were created separating them from bigger counties. Not all of the sources we used for our data were up to date, so that the number of provinces results to be lower in some cases. By looking at the summary table created witht the R studio **car** package, the assumption of great variance in both socioeconomic and climate variables across Italy is confirmed.

Data Description

Statistic

N

Mean

St. Dev.

Min

Max

unemployment_rate

107

8.05

3.79

2.13

18.97

population

110

540,306.80

582,856.20

57,329

3,997,465

gini_index

107

0.30

0.03

0.27

0.36

st_suicide_rate

108

7.49

2.68

0.00

19.19

avg_precipitations

103

870.41

145.51

449.00

1,209.30

avg_temperature

103

13.48

3.20

1.80

19.15

avg_temp_range

103

8.95

1.16

6.40

12.30

gdp_per_capita

110

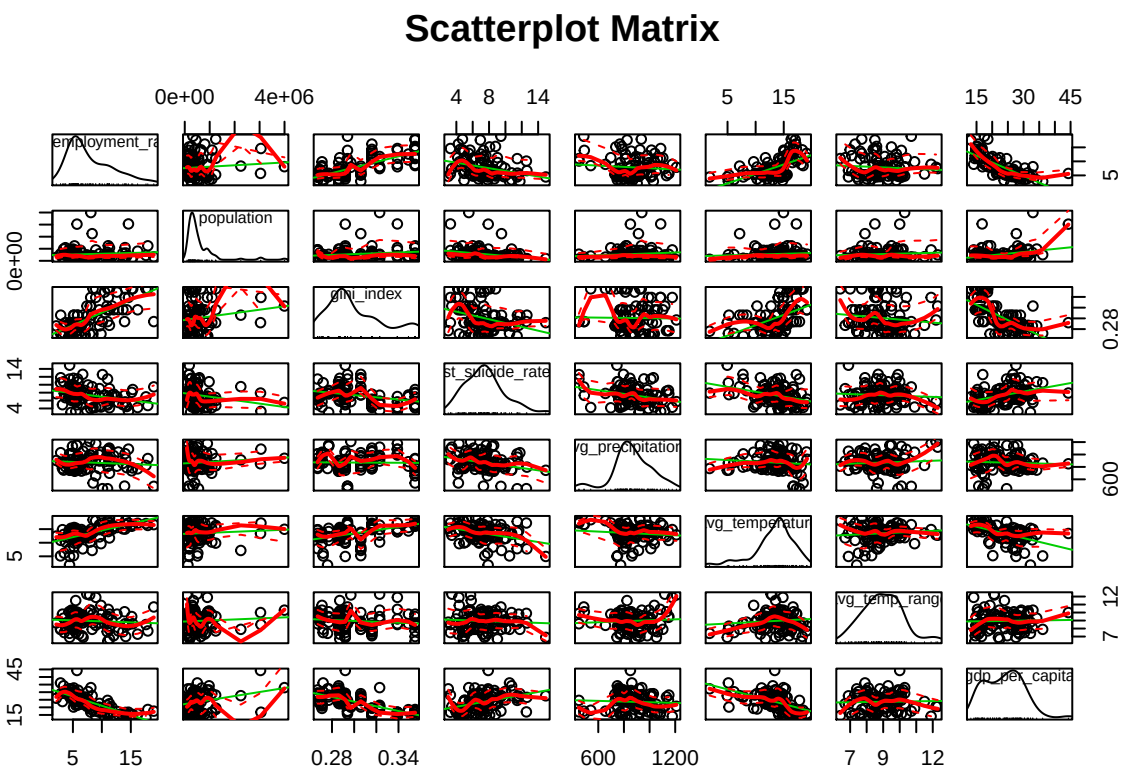
23.33

6.10

11.00

44.30

In order to look at the variables' distribution and find out whether we need to normalize them, we created a scatterplot matrix. From the table below, variables employment, population, gini index, average temperature and GDP per capita seem to be skewed. In order to get more appropriate results, all of the variables will be log-transformed.



Proposed Model:

After having looked at the data and at their characteristics, it can be proceeded to the model estimation. As explained above, no panel data analysis will be conducted, instead we opted for an OLS estimation.

The initial model we want to estimate will therefore be the following:

$$LogStSuicideRate = \alpha + \beta_1 LogGDPPerCapita + \beta_2 LogUnemploymentRate + \beta_3 LogPopulation + \beta_4 LogGiniIndex + \beta_5 LogA$$

We use log-transformed varibales, in order to correct for skewed distributions. The model should show if our initial hypothesis that standardized suicide rates of each one of the 110 Italian provinces is affected by climate and environmental factors, while other socioeconomic indicators are used as control variables.

Estimation of the First Model

Dependent variable:

log(st_suicide_rate)

log(gdp_per_capita)

0.699***

(0.211)

log(unemployment_rate)

0.086

(0.106)

log(population)

-0.174***

(0.043)

log(gini_index)

0.173

(0.487)

log(avg_temperature)

-0.072

(0.097)

log(avg_temp_range)

-0.066

(0.211)

log(avg_precipitations)

-0.435***

(0.147)

Constant

5.312***

(1.415)

Observations

103

R2

0.372

Adjusted R2

0.326

Residual Std. Error

0.264 (df = 95)

F Statistic

8.041*** (df = 7; 95)

Note:

p<0.1; ***p<0.05;*** p<0.01

A second, simplified, model:

$$SuicideRate = \alpha + \beta_1 + \beta_2 UnemploymentRate + \beta_3 population + \beta_4 average_t emperature + \beta_5 average_p recipitations + e$$

Estimation of the Second Model

Dependent variable:

st_suicide_rate

gdp_per_capita

0.199***

(0.057)

unemployment_rate

0.115

(0.087)

log(population)

-1.265***

(0.299)

avg_temperature

-0.115

(0.073)

avg_precipitations

-0.004***

(0.001)

Constant

23.081***

(3.518)

Observations

103

R2

0.381

Adjusted R2

0.349

Residual Std. Error

1.843 (df = 97)

F Statistic

11.950*** (df = 5; 97)

Note:

p<0.1; ***p<0.05;*** p<0.01

Models Comparison

Dependent variable:

log(st_suicide_rate)

st_suicide_rate

(1)

(2)

log(gdp_per_capita)

0.699***

(0.211)

log(unemployment_rate)

0.086

(0.106)

gdp_per_capita

0.199***

(0.057)

unemployment_rate

0.115

(0.087)

log(population)

-0.174***

-1.265***

(0.043)

(0.299)

log(gini_index)

	0.173	
	(0.487)	
log(avg_temperature)		
	-0.072	
	(0.097)	
log(avg_temp_range)		
	-0.066	
	(0.211)	
log(avg_precipitations)		
	-0.435***	
	(0.147)	
avg_temperature		
	-0.115	
	(0.073)	
avg_precipitations		
	-0.004***	
	(0.001)	
Constant		
5.312***		23.081***
(1.415)		(3.518)
Observations		
103		103
R2		
0.372		0.381
Adjusted R2		
0.326		0.349
Residual Std. Error		
0.264 (df = 95)		1.843 (df = 97)
F Statistic		
8.041*** (df = 7; 95)		

11.950*** (df = 5; 97)

Note:

p<0.1; ***p<0.05;*** p<0.01

Table of Coefficients

(Intercept)

23.081

gdp_per_capita

0.199

unemployment_rate

0.115

log(population)

-1.265

avg_temperature

-0.115

avg_precipitations

-0.004

Table of Coefficients

2.5 %

97.5 %

(Intercept)

16.098

30.064

gdp_per_capita

0.086

0.311

unemployment_rate

-0.057

0.287

log(population)

-1.858

-0.672

avg_temperature

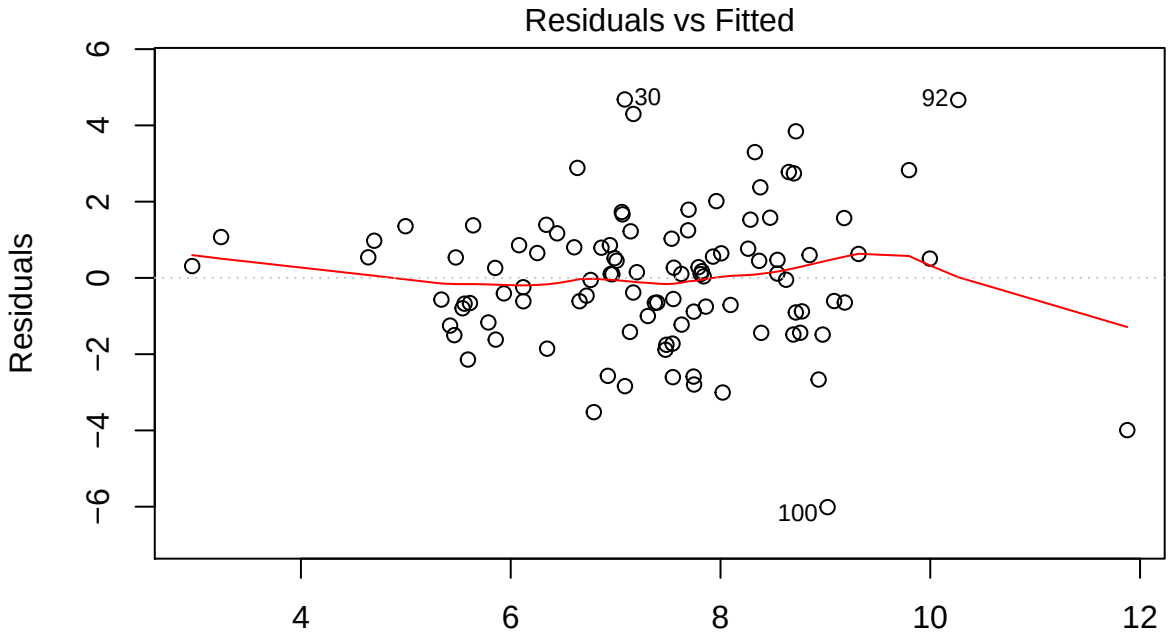
-0.260

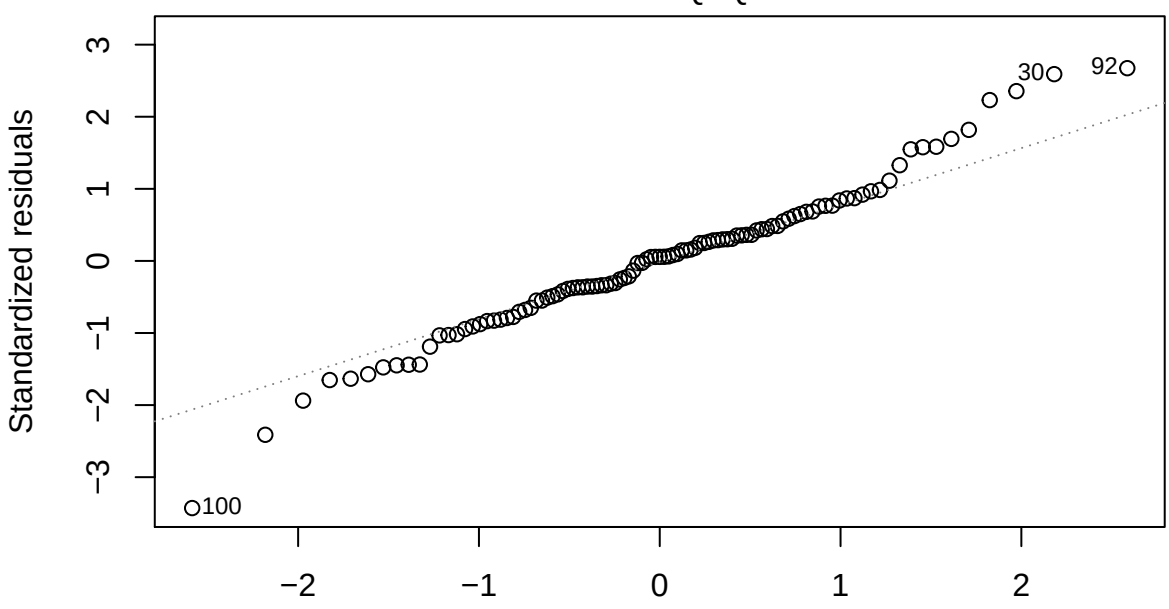
0.029

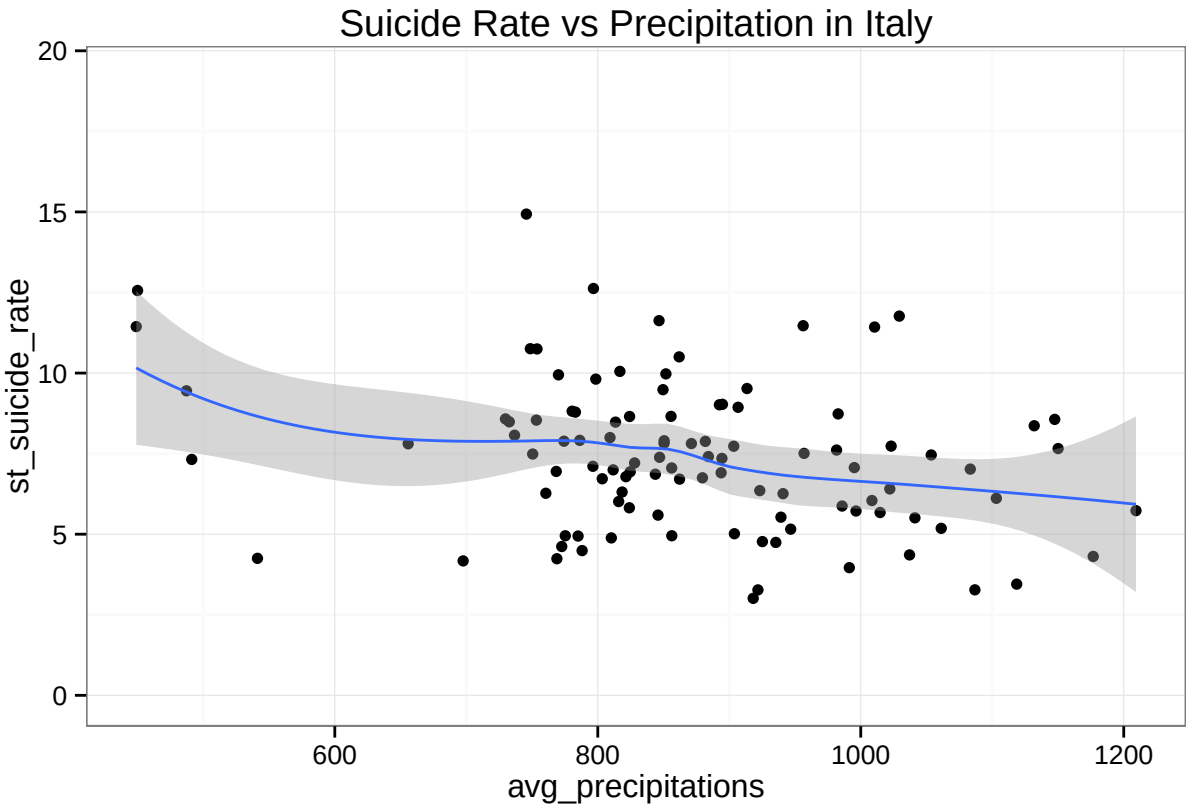
avg_precipitations

-0.007

-0.001


Residuals vs Fitted
Residuals
Fitted values
lm(st_suicide_rate ~ gdp_per_capita + unemployment_rate + log(population) + ...
30
92
100


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(st_suicide_rate ~ gdp_per_capita + unemployment_rate + log(population) + ...
30
92
100



Main findings and discussion:

Conclusion:

References:

Berry, Helen Louise, Kathryn Bowen, and Tord Kjellstrom. 2009. "Climate Change and Mental Health: a Causal Pathways Framework."

Bouchama A., Mohamed G., Dehbi M. 2007. "Prognostic Factors in Heat Wave Related Deaths: a Meta-Analysis." *Arch Intern Med.*, no. 20: 2170–76.

Gandrud, Christopher. 2014. *repmis: A Collection of Miscellaneous Tools for Reproducible Research with R.* http://CRAN.R-project.org/package=repmis.

ISTAT. 2012. "I Suicidi in Italia: Tendenze E Confronti, Come Usare Le Statistiche."

Magnusson, A. 2007. "An Overview of Epidemiological Studies on Seasonal Affective Disorder." *Acta Psychiatria Scand.*, no. 101: 176–84.

OECD. 2013. "Health at a Glance 2013: OECD." http://dx.doi.org/10.1787/health_glance-2013-en.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

WHO. 2014. "Preventing Suicide: a Global Imperative."

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (2014).

Wickham, Hadley, and Winston Chang. 2014. *ggplot2: An Implementation of the Grammar of Graphics.* http://CRAN.R-project.org/package=ggplot2.