

world happiness - kaggle dataset

patrick mcdevitt (& collaboration with preeti swaminathan)

22 may 2017

Happiness - Regression Project -

Intro

The following **Context** and **Content** is from : <https://www.kaggle.com/unsdsn/world-happiness>

Context

The World Happiness Report is a landmark survey of the state of global happiness. The first report was published in 2012, the second in 2013, and the third in 2015. The World Happiness Report 2016 Update, which ranks 156 countries by their happiness levels, was released today in Rome in advance of **UN World Happiness Day, March 20th**.

Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness. They reflect a new worldwide demand for more attention to happiness as a criteria for government policy.

Content

The happiness scores and rankings use data from the Gallup World Poll. The scores are based on answers to the main life evaluation question asked in the poll. This question, known as the Cantril ladder, asks respondents to think of a ladder with the best possible life for them being a 10 and the worst possible life being a 0 and to rate their own current lives on that scale. The scores are from nationally representative samples for the years 2013-2015 and use the Gallup weights to make the estimates representative. The columns following the happiness score estimate the extent to which each of six factors :

- * economic production,
- * social support,
- * life expectancy,
- * freedom,

- * absence of corruption, and
- * generosity
- contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world’s lowest national averages for each of the six factors.

Objective

Within the context of **World Happiness Day** explore the data available from the the United Nations survey and idneitfy the characteristics that are associated with higher **happiness** scores

Descriptive statistics

Analysis

Interpretation

Conclusion

___Setup some data directories

Remove extreme outliers from each column of numerical data

```
-----
region_code
-----
```

Make first data visualization plot for each column of dataset

- numeric data columns
 - histogram with some summary statistics
 - x-y plot vs. dependent variable (in this case = happiness score)
 - x-y plot of residuals of each independent variable linear fit to happiness_score
- categorical columns
 - histogram of counts of each different value in category
 - boxplots within that column of data for view of variability by factor

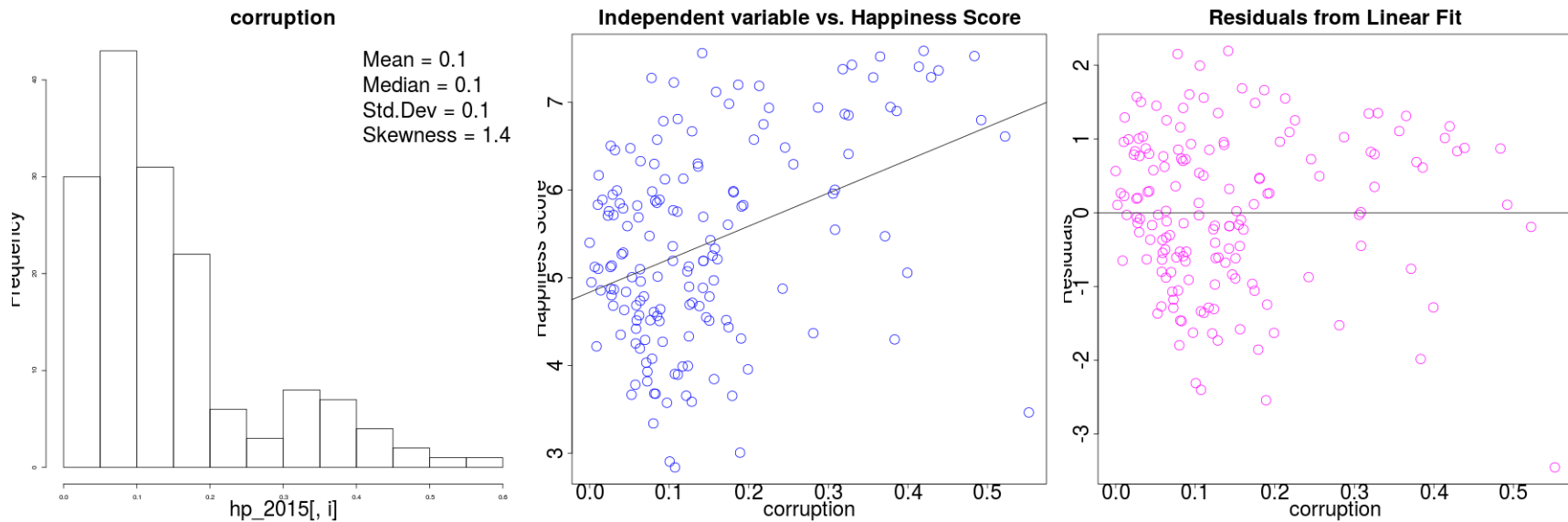


Figure 1: **First Data Views - histogram, single parameter linear fit, fit residuals**

Observations from first data views :

1 - each column is well-behaved ... in the sense that there are not any significant outliers, and - gdp_per_capita

pdf

2

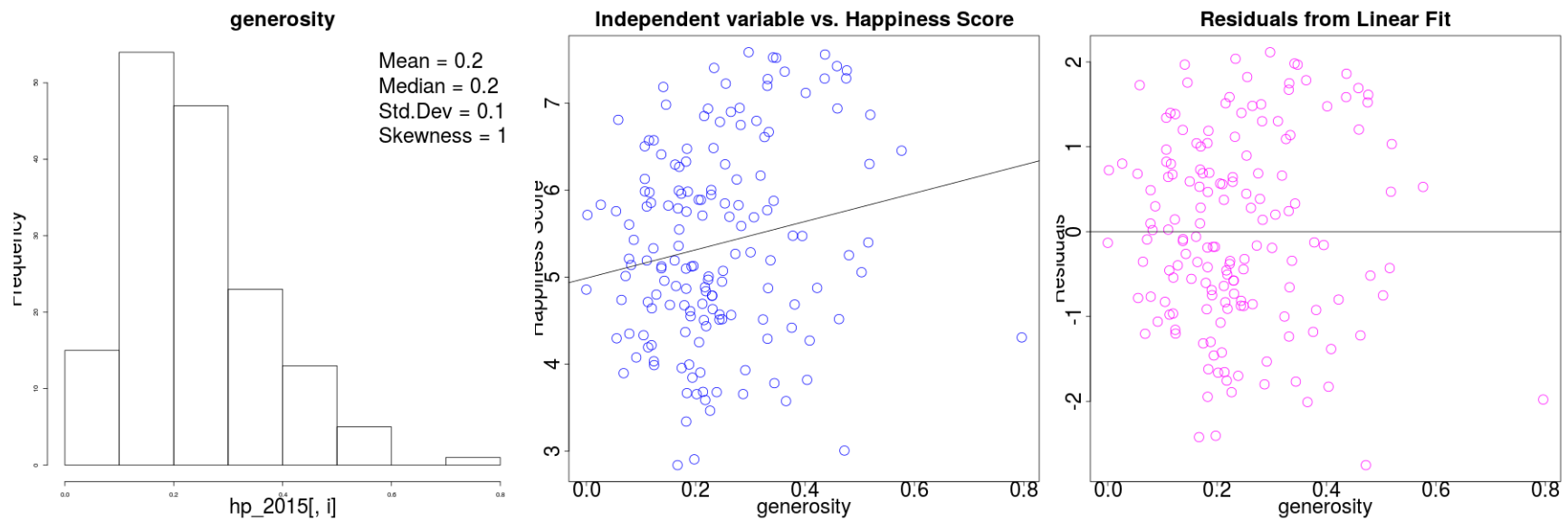


Figure 2: **First Data Views - histogram, single parameter linear fit, fit residuals**

pdf
2