**W203: Exploring and Analyzing Data Section 4 – Amitava Das**

**Fall 2014**

**Lab 1**

Part 1: Multiple Choice

1. e

2. e

3. d

4. b

5. b

6. b

7. c

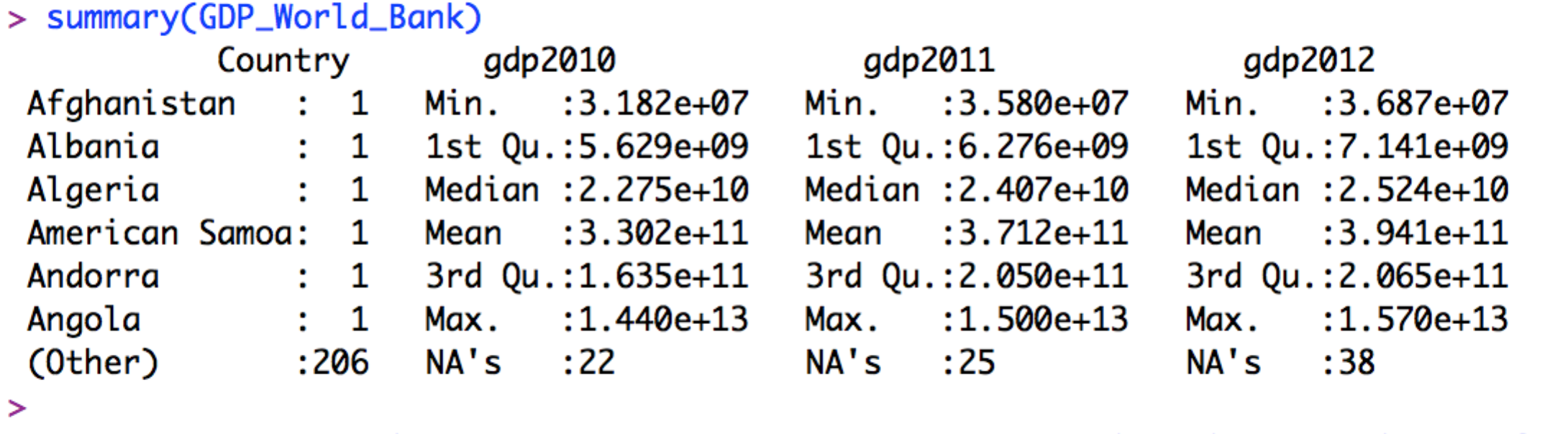
8. d

9. b

Part 2: Data Analysis

1. Variable manipulation

a. The data was loaded from the provided file into a dataframe GDP\_World\_Bank in a locally referenced working directory. The summary of the dataframe is provided below:

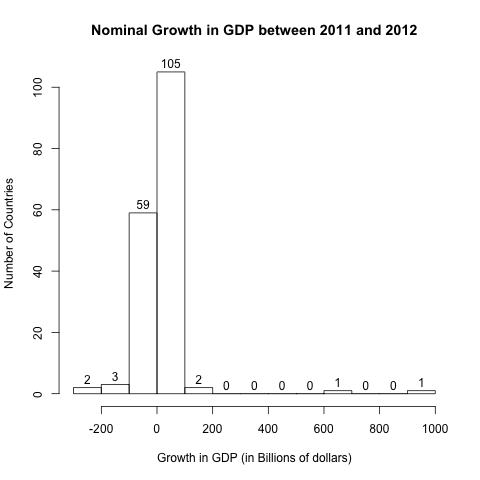


The nominal increase in gdp from 2011 to 2012 was captured in a variable gdp\_growth and the average was calculated after removing the NA values



i.e, the average nominal growth in gdp between 2011 and 2012 was approximately USD 7B (7,172,376,796)

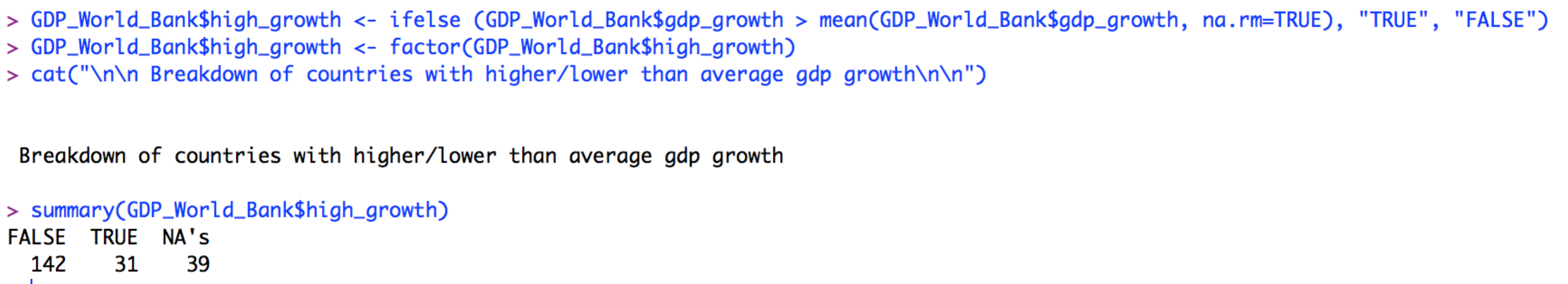
b. Histogram



The distribution is positively skewed, with a tail of values to the right, and leptokurtic (high peaks). 2 data points seem to be outliers with more than 700 and 900 billion in gdp growth, which will definitely skew the mean.

A histogram drawn by scaling the variable to its logarithmic values yielded a more representative normal distribution

c. The new variable high\_growth was set to “TRUE” where the value of gdp\_growth was greater than the mean of the variable gdp\_growth.



As we can see form the output above, 31 countries had higher than average nominal growth in GDP, while 142 countries had lower than average growth.

This can be expected after our review of the distribution of the gdp\_growth variable via the histogram generated in part b, which showed a very positively skewed distribution. The histogram indicated that most of the countries had smaller increases in nominal gdp, and that 2 data points were outliers and would distort the mean

2. Data Source: Institute for Economics and Peace

A global non-profit research firm, the IEP conducts and disseminates research related to business, peace and economic development. The data can be requested and downloaded at:

<http://economicsandpeace.org/research/iep-indices-data/global-peace-index>

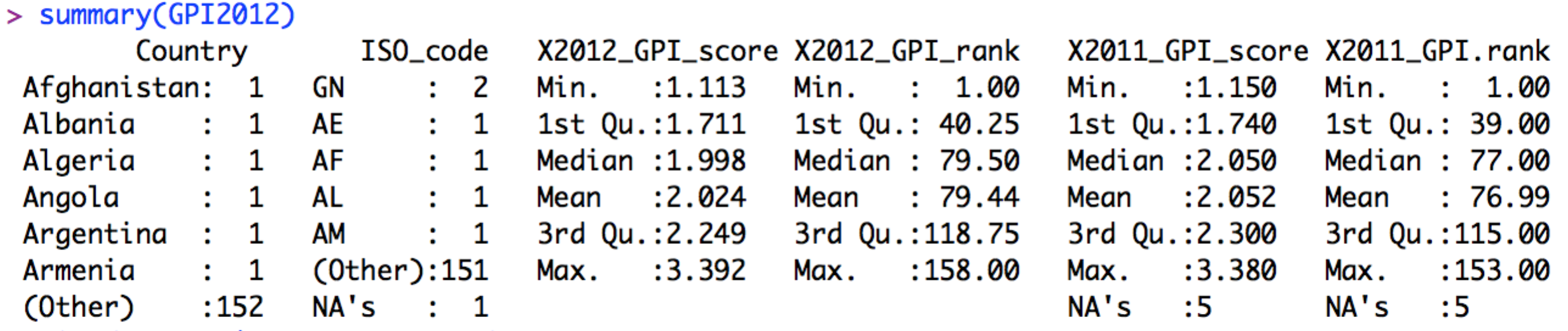
Since I needed data for an earlier year, I was able to download the data for the year 2012 from:

<http://www.theguardian.com/news/datablog/2012/jun/12/global-peace-index-2012>

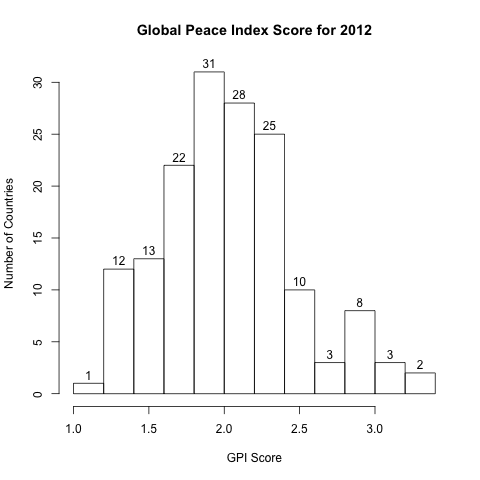
The data needed only the minimal amount of cleansing:

* Removed columns unnecessary for the purpose of this assignment
* Changed the names of some countries to align better with the existing data set – removal of apostrophes and quotes

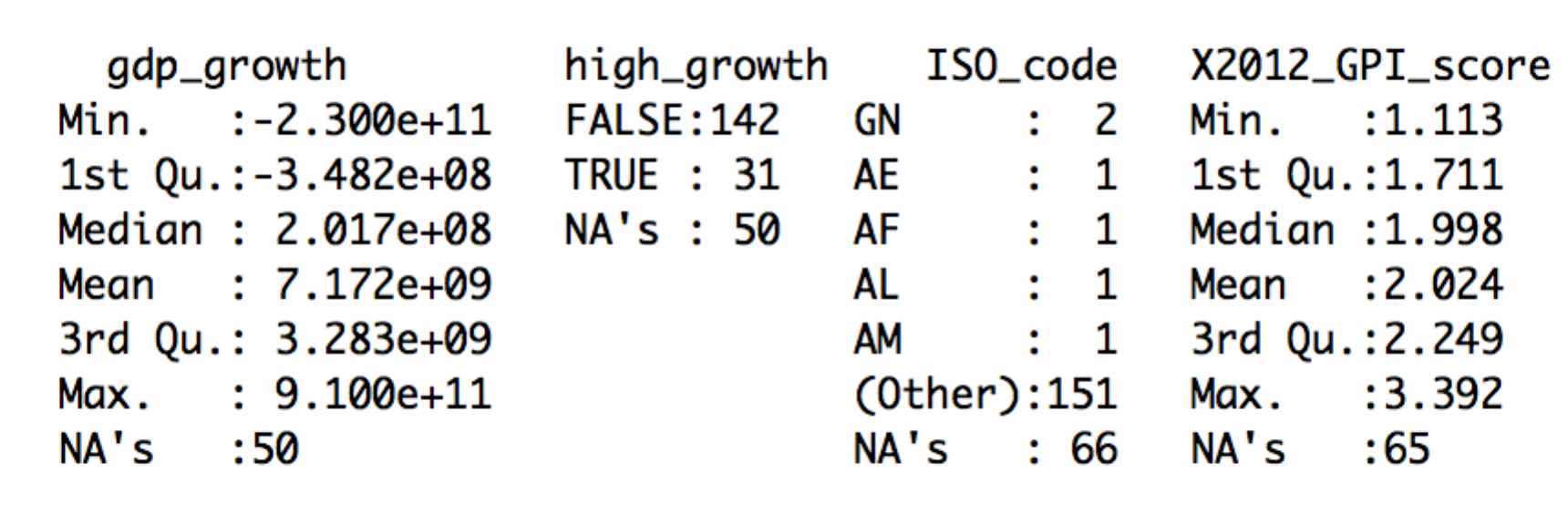
A quick summary:



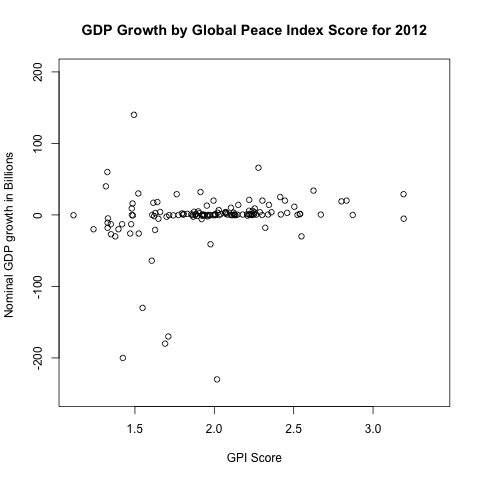
A brief histogram to review the general distribution of the data:



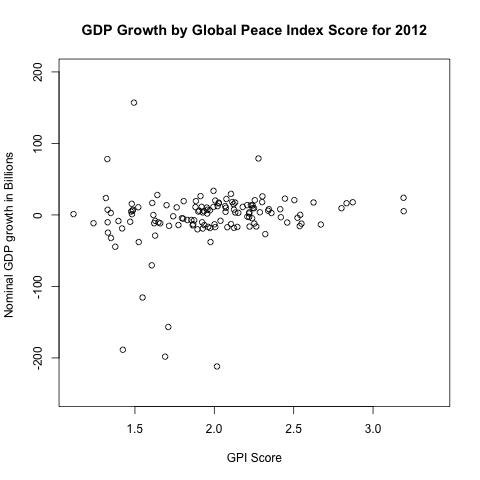
The data was merged using the variable Country, a snapshot of the summary of the relevant columns:



Finally, a scatterplot was generated to view the relationship between the Global Peace Index (GPI) and the nominal growth in GDP for the year 2012



The data was plotted again with the jitter function to reduce the overplotting



The data seems to indicate that the level of peace in a country does not have a significant correlation with a nominal growth in GDP.