Assignment 1

Alex Sacharow & Torben Klausa

October 07, 2016

Contents

1 Introduction

In this assignment we draw two data sets from the GitHub repository fivethirtyeight and apply some descriptive statistics on them.

2 Data Set 1: Police Killings

The first data set we choose is about police killings in the US which was gathered by the Guardian. Information on the structure of the raw data set can be accessed on GitHub: fivethirtyeight. FiveThirtyEight wrote themselves about it on their website and the corresponding Guardian story is available on The counted. What fallows are our own descriptive statistics of the data set.
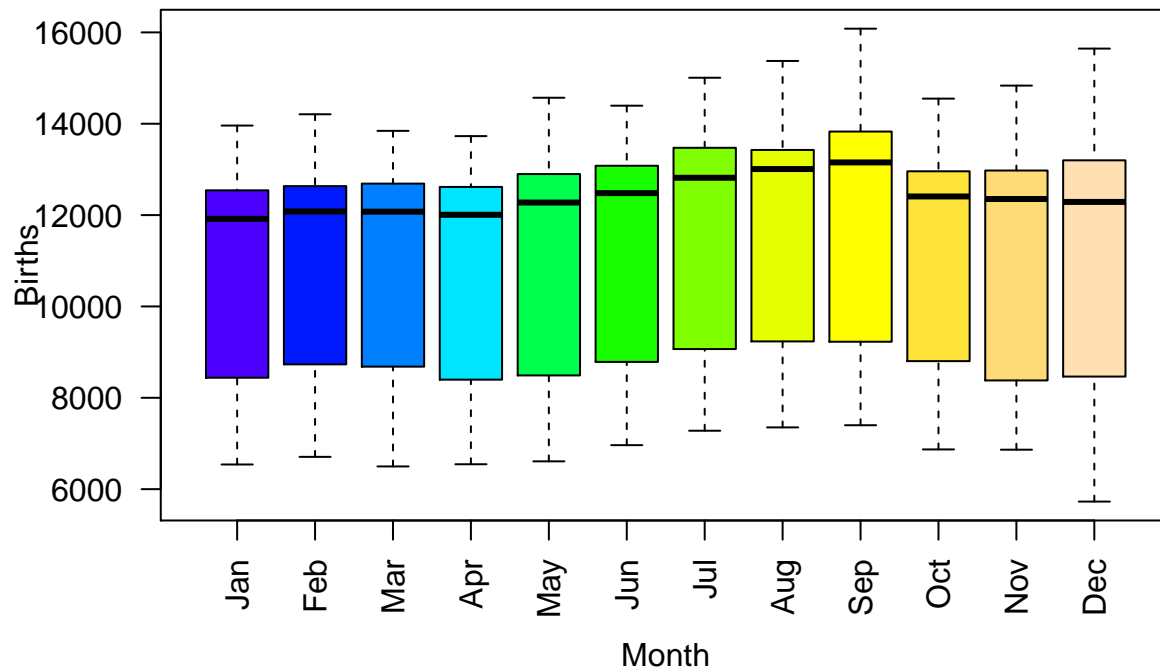
3 Data Set 2: US Birth Statistics

The second data set, again drawn from GitHub: fivethirtyeight, contains data on daily births in the U.S. for the years 2000 to 2014, as provided by the Social Security Administration.

The average number of births per day in this time period is 1.1350068×10^4 with a standard deviation of 2325.8210494.
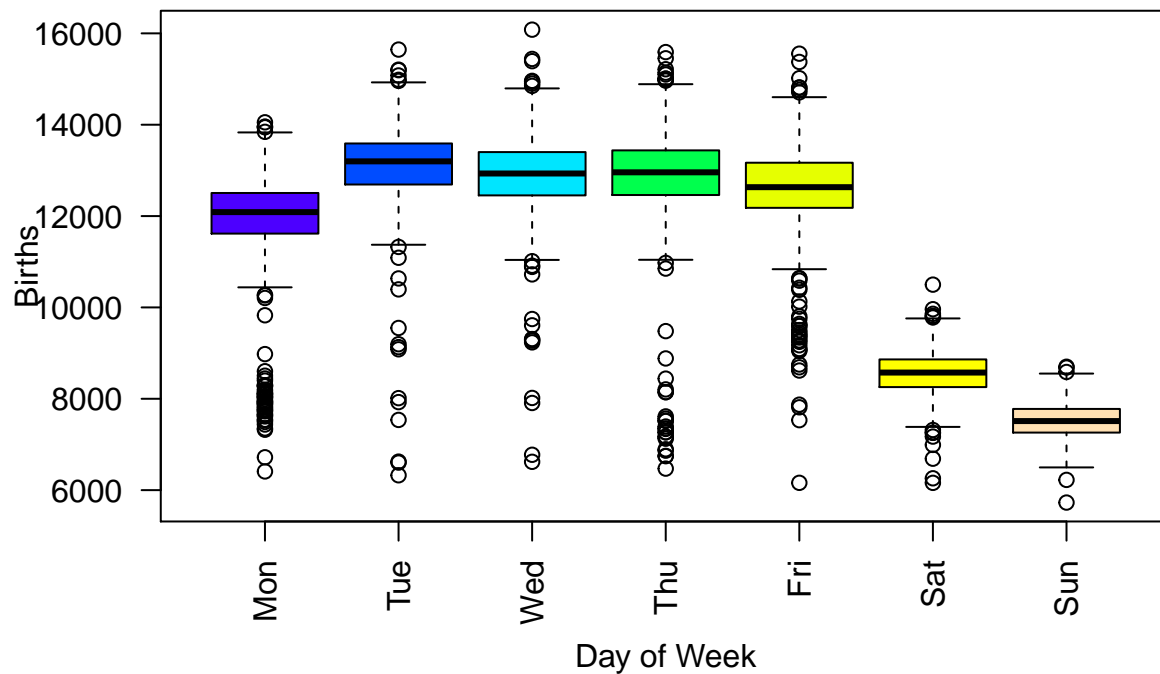
3.1 Births per Day in Different Months

However, we see some differences between the months, hinting to the fact that many people probably use the cold winter months to engage in baby-generating activities:

3.2 Births on Single Days of the Week

Another interesting view is generated by plotting the number of births on the respective days of the week. What we can see here is that the weekend is used by a much lower amount of people to give birth to their babies:

3.3 Births, Days of Week, and Months

Finally, a more comprehensive view is offered by a heat map, showing that compared to other weekdays and months, on Tuesdays in July and from Tuesdays to Fridays in August most babies in the U.S. are born.