**Option 1 Assignment**

ZHANG Chenlin

1155151471

**My chosen website for performing the scraping:**

<https://www.basketball-reference.com/draft/BAA_1947.html>

**README**

I scraped the data from 1947 to 2019 on the NBA website, using ParseHub, as you can see below.

手机屏幕截图

描述已自动生成

As suggested in the article (<http://savvastjortjoglou.com/nba-draft-part01-scraping.html>), I cleaned the data from 1966 to 2014 with OpenRefine.

手机屏幕截图

描述已自动生成

Basically, I performed 4 actions cleaning the original CSV data in OpenRefine:

1. Deleting data from 1947 to 1965 and from 2015 to 2019;
2. Renaming the columns to align with the website;
3. Replacing the “NULL” value with “0”, using GREL language *if(value==null,'0',value)*, with the help of instructions on <https://github.com/OpenRefine/OpenRefine/wiki/GREL-String-Functions>
4. Transforming data type: text to number; text to date.