<h1 style="text-align:center">An Analysis of Official Development Finance from China</h1>

**I.   Introduction**

   a. Data Source

The AidData Lab at the College of William and Mary has led the work to probe into the mystery of Chinese foreign investment, publishing several versions of AidData's Global Chinese Official Finance Dataset (2000-2014) with the link attached below. In this study, we will try to learn more about the behaviors of official Chinese foreign investment using data wrangling, text mining, and machine learning techniques.

https://www.aiddata.org/data/chinese-global-official-finance-dataset

   b. Snapshot of projects' value by years, including the total, average, min, and max

```
   Year `Number of Proj… `Total Value of… `Average Value …
  <dbl>            <int>            <dbl>            <dbl>
1  2000              101      2337247008.        32461764.
2  2001              135      5455489633.        50050364.
3  2002              134      4376162094.        46064864.
4  2003              181      5525240301.        46824070.
5  2004              181      4520528562.        40004678.
6  2005              242      8045518295.        49663693.
7  2006              314     13831559135.        73965557.
8  2007              327     15695022812.        75821366.
9  2008              278      8530799541.        55756860.
10 2009              383     64117820118.       276369914.
11 2010              342     24826825668.       134928400.
12 2011              390     44008421953.       202803788.
13 2012              336     37135252439.       200731094.
14 2013              342     27958099383.       158852837.
15 2014              327     30997892020.       190171117.
# … with 2 more variables: `Minimum Value of Project` <dbl>, `Maximum
```


Graph 1: Total Value of Projects by Years

Graph 2: Total Value of Projects by Years
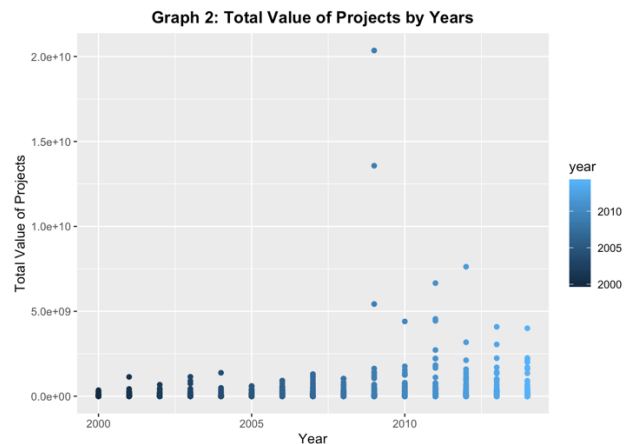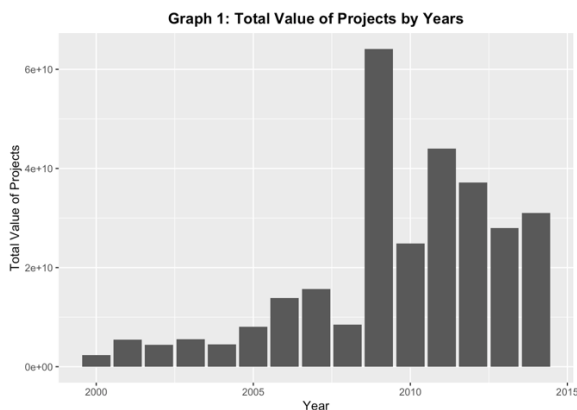
The data show an overall increasing pattern for the total value of projects by years, with a few setbacks along the way. Graph 1 displays a clearer picture of the overall pattern as we can see the highest flow happened in 2009, which followed the sharp recession in 2008 and followed by a sharp recession in 2010. Graph 2 indicates that

there are some outliers among projects in years, such as 2009, which helps to explain the reason why there is a peak in 2009 for the total value of projects.

c. Total projects' value by 2014 USD by regions

Total Value of Projects by Regions


```
## # A tibble: 6 x 2
##   Region                          `Total Value of Projects`
##   <chr>                                             <dbl>
## 1 Asia                                         95822989635.
## 2 Africa                                       92713211229.
## 3 Central and Eastern Europe                   55569462403.
## 4 Latin America and the Caribbean              50403066261.
## 5 The Pacific                                   2337839964.
## 6 Middle East                                    515309469.
```

From the data, we can see that Asia receives the highest value of projects. It is evident to see in the pie chart that Asia and Africa have an almost equal, significant share of the total value of projects, while the Pacific and the Middle East are regions with the least share of value.

II. Texting Mining


```
## # A tibble: 19,002 x 2
##     word           n
##     <chr>        <int>
##  1 china         4929
##  2 chinese       3899
##  3 project       3350
##  4 million       3079
##  5 government    1759
##  6 signed        1702
##  7 loan          1660
##  8 construction  1649
##  9 usd           1483
## 10 agreement     1380
## # … with 18,992 more rows
```

There is a snapshot of topics mentioned in the description of all projects. Keywords, such as construction, equipment, development, bank, medical, ambassador, and government, are worth paying attention to.

III. Predictive Models

a. Predictive Question

Whether a project is ODA or not?

b. Term Explanation

ODA means Official Development Assistance, the major type of official development finance, as opposed to Other Official Flows (OOF). The Global Chinese Official Finance Dataset includes a dataset on China's Official Development Assistance (ODA)-like projects and a dataset on China's Other Official Flows (OOF)-like projects. The reason projects are divided into ODA-like, OOF-like is that ODA and OOF are OECD countries' definitions. ODAs are traditionally referred to as aids, whereas OOFs are just investments. Since Chinese aid and investment do not use such a standard, researchers at AidData have to discern the categories by themselves so that the data could be comparable to western investment and aid. There are about 4000 plus ODA projects and 1100 OOF projects covering the time span of 2000 to 2014.

c. Predictive Models

Rationale: Since the outcome variable in this predictive model is binary, this study chooses the following models to run.

Models: logit model, penalized logit model, linear probability model (LPM), robust LPM, LDA, QDA, PLSDA, nearest shrunken centroids, boosting, random forest, neural network model, and MARS model

Steps: For each model, it follows four steps: 1) Create new variables and split data into training and testing data; 2) Optimize the model through training data; 3) Predict the model using testing data; 4) Evaluate the model by comparing training and testing data

d. Conclusion

According to the rank of the overall prediction accuracy, the penalized logit model and boosting model perform the best in predicting whether the project is an ODA or not with the highest accuracy. Meanwhile notably, the neural network model has the highest balanced accuracy among all models. With a relatively high level of prediction accuracy, this study will help categorize Chinese foreign investment in the future. As China does not classify its foreign investment by the OECD standard, it is very hard to directly compare its investment efficiency with the aid efficiency of Western countries or that of the World Bank. Therefore, this study would, to a certain degree, help address this problem, and provide a benchmark for researchers to compare Chinese and Western foreign investment in the right categories.