## Basic Firm Information

1. In Brandt et al (2014), it is said the Chinese Industry Survey data reports how many plants/establishments a firm has. Is this the variable cyhddwshj (产业活动单位数合计) in the data?
2. The data files are organized by year. Do we already have a panel id in the data to track firms over time?

## 2. Location

1. The location variable is the location of manufacturing plant or the location of the offices, or are these the same? Is there any way we can check this?

## Pollution and Pollution Control

1. Is the source of pollution datafrom direct monitor at the end of a plant’s pipeline?

Jack: 企业在其所在工厂有专门检测的仪器计算具体的排放量，然后加总到企业层面。排污的数据是工厂层面的数据，是每一家工厂的信息

1. Almost all the pollution and control information are missing in 2011-2014. Is there some issue with the data for these years? The variables we looked at are: 工业废水排放量吨, 化学需氧量排放量千克, 氨氮排放量千克, 工业废气排放总量万标立方米, 氮氧化物排放量千克, 二氧化硫排放量千克, 烟尘排放量千克, 工业粉尘排放量千克, 废水治理设施数套, 废水治理设施处理能力吨日, 工业废水处理量吨, 废气治理设施数套, 废气治理设施处理能力标立方米时, 氨氮去除量千克, 化学需氧量去除量千克, 氮氧化物去除量千克, 其中脱硫设施数套, 其中脱硫设施脱硫能力千克时, 二氧化硫去除量千克, 其中当年新增设施去除的千克, 烟尘去除量千克, 工业粉尘去除量千克

这些变量信息都有的，然后数据的格式是string，你需要destring才行，否则直接统计的话，就会显示全是0的情况

1. For 1998-2010, does missing pollution data mean not monitored or not polluting?

数据是1998-2014的，信息都是有的

## Energy and Water Use

1. Almost all the energy and water information are missing in 2011-2014. Is there some issue with the data for these years? The variables we looked at are: 煤炭消费总量吨, 其中燃料煤消费量吨, 原料煤消费量吨, 料油消费量不含车船用吨, 其中重油吨, 柴油吨, 洁净燃气消费量万立方米, 工业用水总量吨, 其中新鲜水量吨, 重复用水量吨.

数据也是有的

## 5. Annual Firm Counts

1. Why there are a lot more firms in 2004, 2008 and 2010? Are 2004 and 2008 due to census? Does 2013 come from census? We want to know the selection criteria into the sample because having a consistent count of firms in each province or other level of geographical location is crucial for us to understand how the variation in local environmental regulations affect entry and exit of firms.

是的，2004和2008是census，这两年数据会稍微多些。另外，1998-2010年工企的进入门槛是主营业务超过500万的企业，但由于考虑到经济增长导致企业数量越来越多，2011年开始进入门槛提升到了2000万，因此2010年是最后一年，数量多一些是正常的

## 6. Representativeness of the Sample

1. One challenge of using this data to stduy firms’ location choices in response to geographical variation of environmental regulations is that we can’t directly track firms’ location choice. However, it may be possible that we can infer firms’ location choices through the number of firms in a geographical location. However, this makes us vulnerable to the sample selection issue of the Chinenese industrial firms. Would it be possible to use census data or local year book to check whether firms in the Chinese industrical firms contribute to the majority of the local production.

## 7. How is the Pollution Data Merged with AMS Data

Jack: 污染数据会公布企业的名称，我们现在用的是企业的名称来匹配的。污染数据是环保局重点观测企业的数据，是一个unique的identity，因此基于污染数据，然后匹配工企数据进去的话，应该是1:1的匹配，也就是说这些企业只有一个工厂。如果工业企业数据库里面的企业有多家工厂的话，按理是可以从污染数据里面找到的。中国所谓的plant不会专门叫plant，就是firm和plant都是同一个location，比如很多工业园里面的企业，既是firm又是plant，制造业企业很少会在一个地方设个firm，然后在另一个地方搞了plant。

XYZ: 理论上讲，我们是可以把multi-plant的企业也进行匹配吗？(Representative sample issue)