

```
import pandas as pd

df = pd.read_stata('text_example_final.dta')

journal1 = df[df['source']==u'管理世界']
journal2 = df[df['source']==u'经济研究']

df.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	title	author	organ	source	keyword	summary	fund	year	id
0	2019 中国 信息 经济 学乌 家培 奖评 选公 告		中国信 息经济 学会;	经济研 究		为促进中国 信息经济与 信息管理领 域的青年学 者产生影响 未来30年世 界理论发展 的创新成果, 中国信息...		2019.0	1.0
1	债务 高 企、 风险 集聚 与体 制变 革—— 对发 展型 政府 的反 思与 超越	张晓晶; 刘学良; 王佳;	中国社 会科学 院经济 研究所 中国宏 观经济 稳定课 题组;	经济研 究	宏观杠杆 率;;债务 风险;;发 展型政 府;;体制 变革	本文通过分 析中国债务 的形成机制, 指出体制性 因素是债务 高企的根本 原因。这一 体制性因素 可概括为国...		2019.0	2.0
2	自然 利率 与中 国宏 观政 策选 择	徐忠;贾 彦东;	中国人 民银行;	经济研 究	自然利 率;;潜在 产出;;全 要素生 产率;;宏 观政策	本文讨论的 核心是如何 准确认识和 理解自然利 率,并尝试建 立以自然利 率为基础的 宏观经济分 析和政策决...		2019.0	3.0
3	中国 国有 企业 的独 特作 用:基 于知 识溢 出的 视角	叶静怡; 林佳;张 鹏飞;曹 思未;	北京大 学经济 学院;中 欧国际 工商学 院;北京 师范大 学经济 与管理 学院;	经济研 究	知识溢 出;;基础 研究;;市 场失灵;; 国有企业	国有企业是 政府解决市 场失灵的工具。本文从 知识溢出角 度研究中国 国有企业的 作用。基于 上市企业样 本...		2019.0	4.0

	title	author	organ	source	keyword	summary	fund	year	id
4	国有资本划转养老保险基金与劳动力长期供给	景鹏;郑伟;	西南财经大学保险学院;北京大学经济学院;	经济研究	国有资本;;养老保险;;划转率;;劳动力供给	国有资本划转养老保险基金通过“挤占公共财政收入”和“放松个人预算约束”两条路径影响劳动力供给...		2019.0	5.0

```

def get_keywords(df):
    dic = {}

    for keywords in df['keyword']:
        keywords = keywords.split(';;')
        for keyword in keywords:
            if(keyword == ""):
                continue
            if keyword not in dic:
                dic[keyword] = 0

            dic[keyword] = dic[keyword] + 1

    lis = []
    for keyword in dic.keys():
        lis.append((dic[keyword], keyword))

    lis = sorted(lis)
    lis.reverse()
    return dic, lis

d2017, l2017 = get_keywords(df[df['year']==2017])
d2018, l2018 = get_keywords(df[df['year']==2018])
d2019, l2019 = get_keywords(df[df['year']==2019])

keyword_dic = {}
for _, keyword in get_keywords(df)[1][0:10]:
    keyword_dic[keyword] = 0

# first = 5

# for _, keyword in l2017[0:first]:
#     keyword_dic[keyword] = 0

# for _, keyword in l2018[0:first]:
#     keyword_dic[keyword] = 0

```

```
# for _,keyword in l2019[0:first]:
#     keyword_dic[keyword] = 0
```

```
pd.DataFrame(l2017).head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}
```

```
.dataframe thead th {
    text-align: right;
}
```

	0	1
0	12	经济增长
1	5	政策建议
2	5	创新
3	5	供给侧结构性改革
4	4	融资约束

```
from matplotlib import pyplot as plt
import matplotlib as mpl
import numpy as np

plt.rcParams['font.family']='sans-serif'
plt.rcParams['font.sans-serif']=['SimHei'] #用来正常显示中文标签
plt.rcParams['axes.unicode_minus']=False #用来正常显示负号

dpi = 200
fig = plt.figure(figsize=(1600/dpi, 1600/dpi), dpi=dpi)

for keyword in keyword_dic.keys():
    x = [2017, 2018, 2019]
    print(keyword)
    y = [d2017.get(keyword) or 0, d2018.get(keyword) or 0, d2019.get(keyword) or 0]
    plt.plot(x,y, label=keyword)

plt.legend(loc=[1.1, 0])
plt.xticks([2017, 2018, 2019])
plt.yticks(np.arange(0, 16, 1))
```

```
plt.savefig('p3_1a.png', figsize=(1600/dpi, 1600/dpi), dpi=dpi)
fig.set_facecolor("#FFFFFF")

print(journal1['year'].min(), journal1['year'].max())
print(journal2['year'].min(), journal2['year'].max())
```

经济增长
 供给侧结构性改革
 国有企业
 创新
 全要素生产率
 融资约束
 收入分配
 公司治理
 人力资本
 货币政策
 2017.0 2019.0
 2017.0 2019.0

```
data = get_keywords(df)[1][:20]
dpi = 200
fig = plt.figure(figsize=(1600/dpi, 400/dpi), dpi=dpi)
plt.xticks(rotation=90)
plt.bar([d[1] for d in data], [d[0] for d in data])
fig.set_facecolor("#FFFFFF")
```

