NameData

May 9, 2023

[3]:
```
import pandas as pd, matplotlib.pyplot as plt, sqlite3, numpy as np, seaborn as␣
 ↪sbn
```

The following database is National Baby Name data sourced from Kaggle:

[4]:
```
con= sqlite3.connect("/Users/clairecrain/Documents/Data Stuff/archive (1)/
 ↪database.sqlite")
```

[5]:
```
cur = con.cursor()
```

The following code creates a temporary table of the Top 5 popular baby names Nationally seperated by gender and year:

[10]:
```
cur.execute('Drop Table if exists RankNames;')

cur.execute('Create temporary Table if not exists RankNames as Select * from␣
 ↪(select Name,Gender,Year,Count,row_number() over (partition by year, gender␣
 ↪order by count desc) as NameRank from NationalNames) where NameRank<=5;')
```

[10]:
```
<sqlite3.Cursor at 0x11d045040>
```

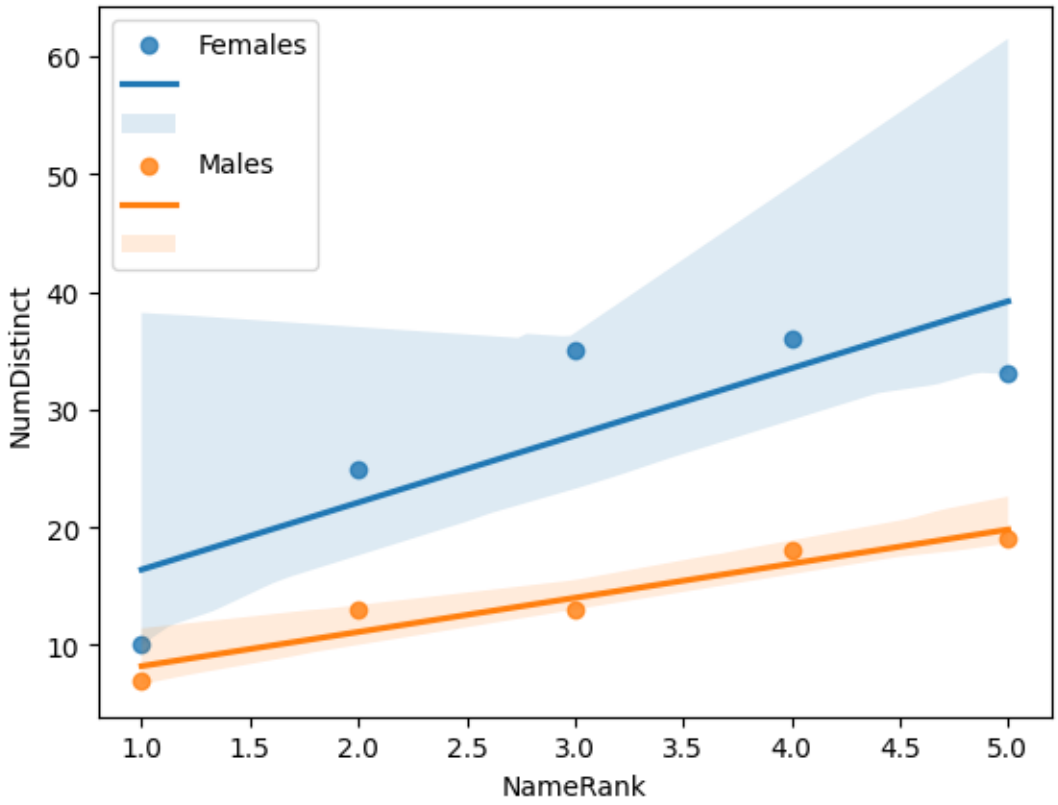
[7]:
```
sqlF = """Select distinct NameRank,Gender,dense_rank()
over (partition by NameRank, Gender order by name) +dense_rank() over
(partition by NameRank, Gender order by name desc) -1 as NumDistinct from␣
 ↪RankNames
where Gender='F';"""

sqlM = """Select distinct NameRank,Gender,dense_rank()
over (partition by NameRank, Gender order by name) +dense_rank() over
(partition by NameRank, Gender order by name desc) -1 as NumDistinct from␣
 ↪RankNames
where Gender='M';"""

data1= pd.read_sql(sqlF, con)
data2= pd.read_sql(sqlM, con)

sbn.regplot(x=data1['NameRank'],y=data1['NumDistinct'])
sbn.regplot(x=data2['NameRank'],y=data2['NumDistinct'])
```

```
leg = plt.legend(loc="upper left",labels=["Females","","","Males","","",""])
```



Although there is 124 years of data some names stay popular year after year. This chart displays the number of distinct names for each ranking over the entire 124 years seperated by gender. As clearly seen above there is signifanctly more variety in the female names than the male for every rank. There seems to be a positive correlation between the amount of variety in the names as the polularity decreases. The correlation appears to be much stronger with male names opposed to female names.

My theory as to why there is so much more female name variation than male is gender roles and disrimination. The social acceptability for names is largely rooted in gender roles. Men are expected to be maculine, strong, providers etc. This results in names that carry that social expectation. Especially in the past 10 years, traditionally boy names used for baby girls has been increasing in popularity; thus giving a wider pool of possible girl names. This is not the same in reverse, traditionally girls names are not being used for baby boys; thus keeping the boy name pool more limited. This is largely in part due to sexism; things attributed to boys are inherently viewed as better than things attributed to girls.

[ ]: