Popular Baby Names - Social Security

BY: JINESH DHRUV

DATASET

- State-specific data on the relative frequency of given names in the population of U.S. births where the individual has a Social Security Number
- Year Range: 1910 – 2016
- Total Records: 5,838,786
- Total Male: 44.2%
- Total Female: 55.8%

https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-data-by-state-and-district-of-#topic=developers_navigation

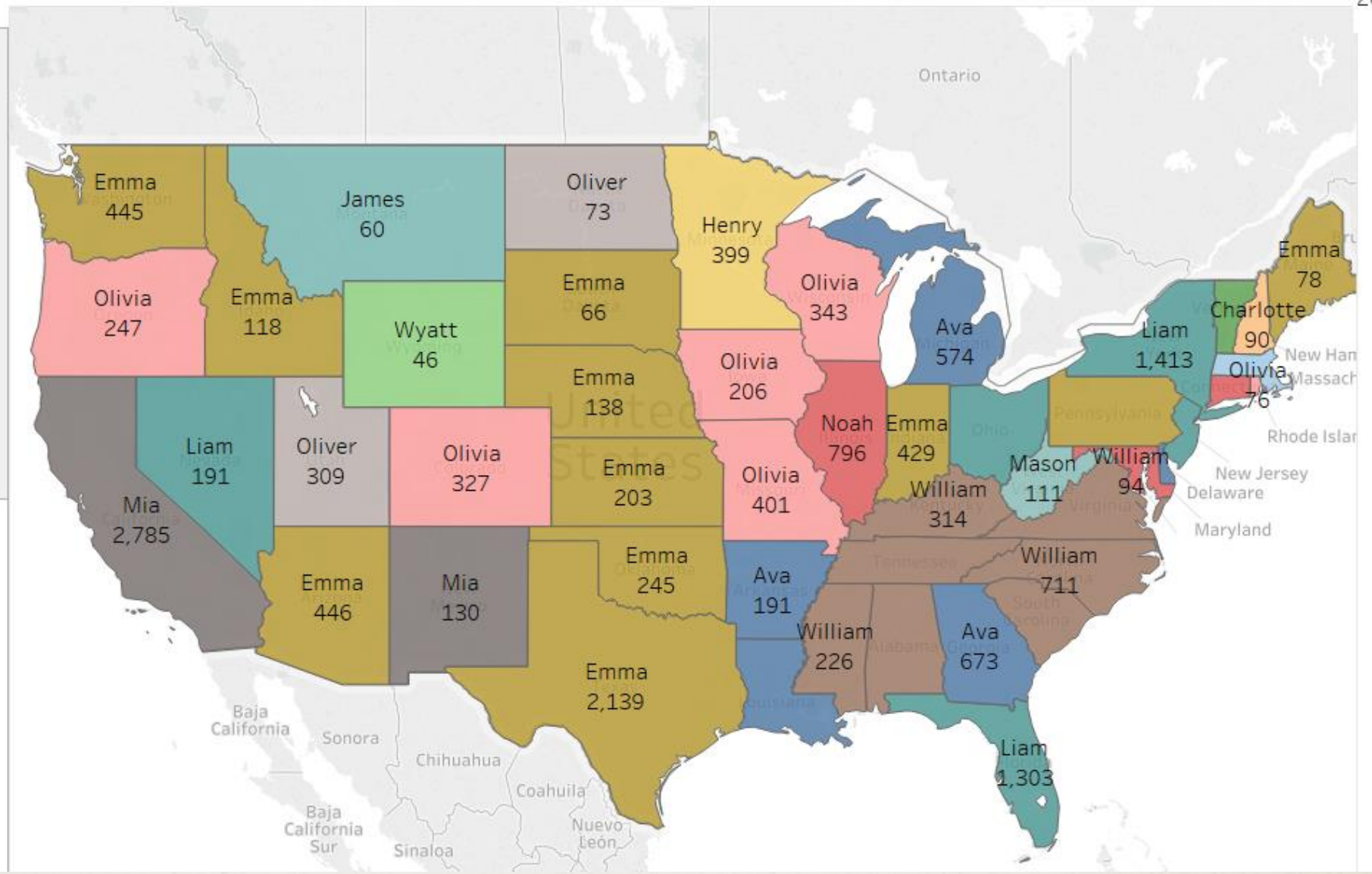
TABLEAU DASHBOARDS

Popular names by state

Most popular name by state

State	Name	Count
AK	Emma	47
AL	William	427
AR	Ava	191
AZ	Emma	446
CA	Mia	2,785
CO	Olivia	327
CT	Noah	222
DC	William	94
DE	Ava	65
FL	Liam	1,303
GA	Ava	673
HI	Noah	78
IA	Olivia	206
ID	Emma	118
IL	Noah	796
IN	Emma	429
KS	Emma	203
KY	William	314
LA	Ava	297
MA	Benjamin	487
MD	Noah	334
ME	Emma	78
MI	Ava	574
MN	Henry	399
MO	Olivia	401
MS	William	226
MT	James	60
NC	William	711

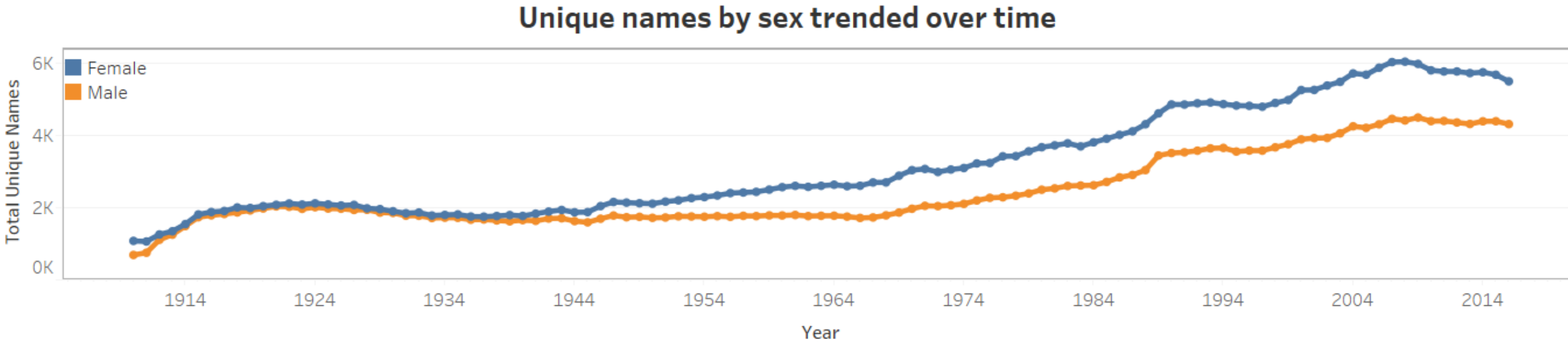
Top names by state from 1910 - 2016

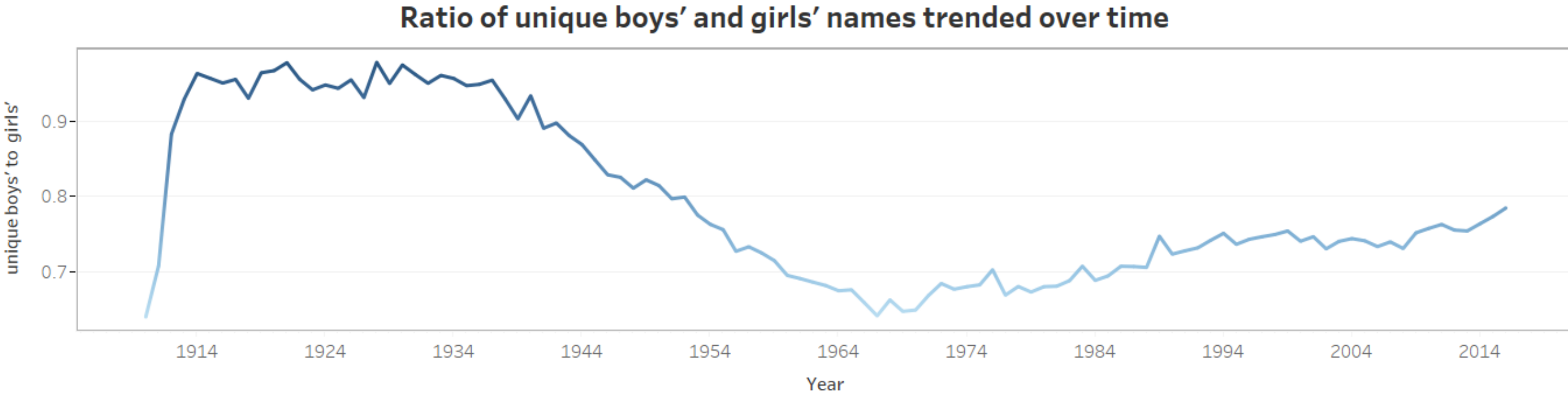


Year

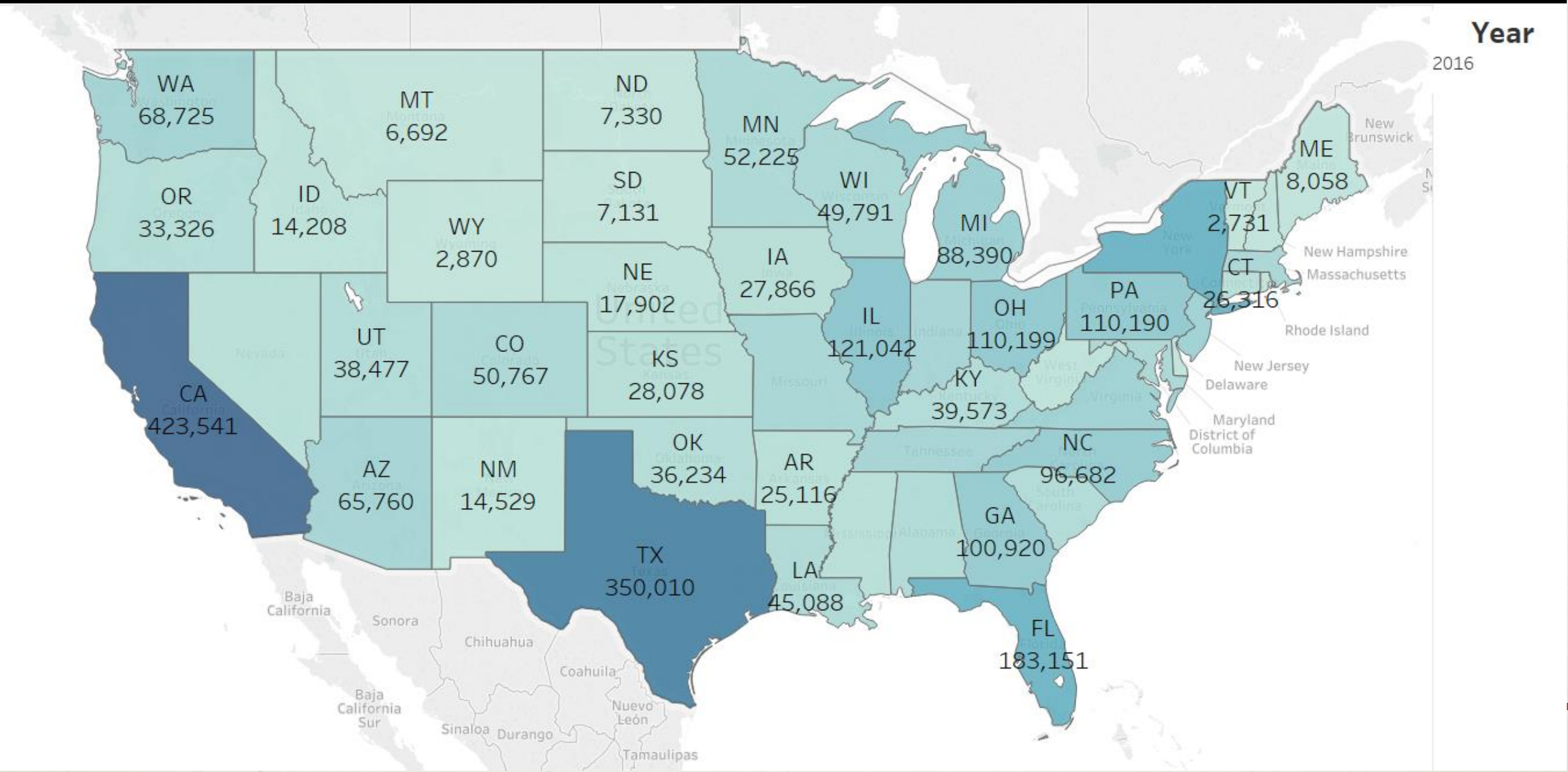
2016

Time series analysis of names by gender



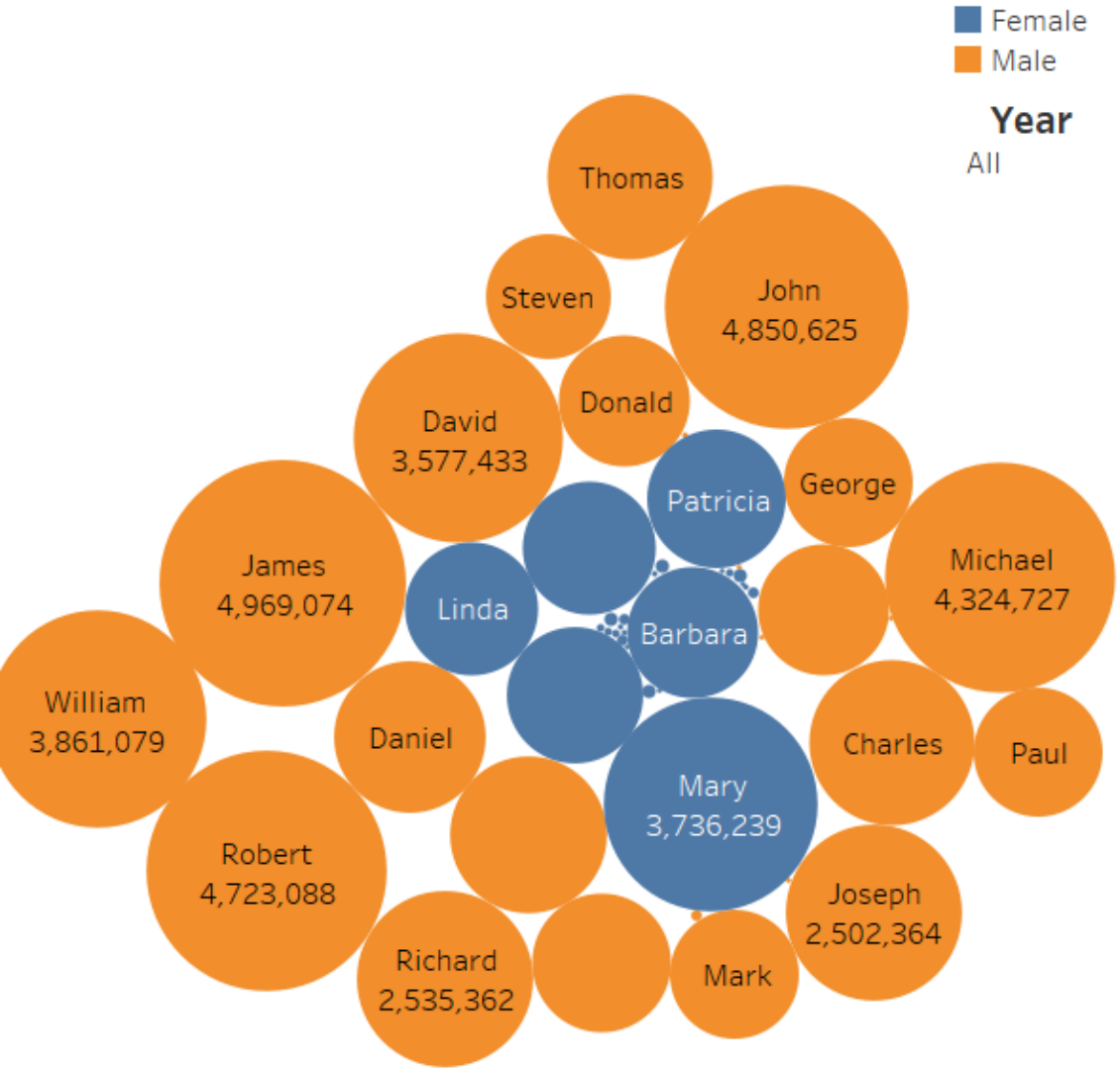


Total names by state and sex

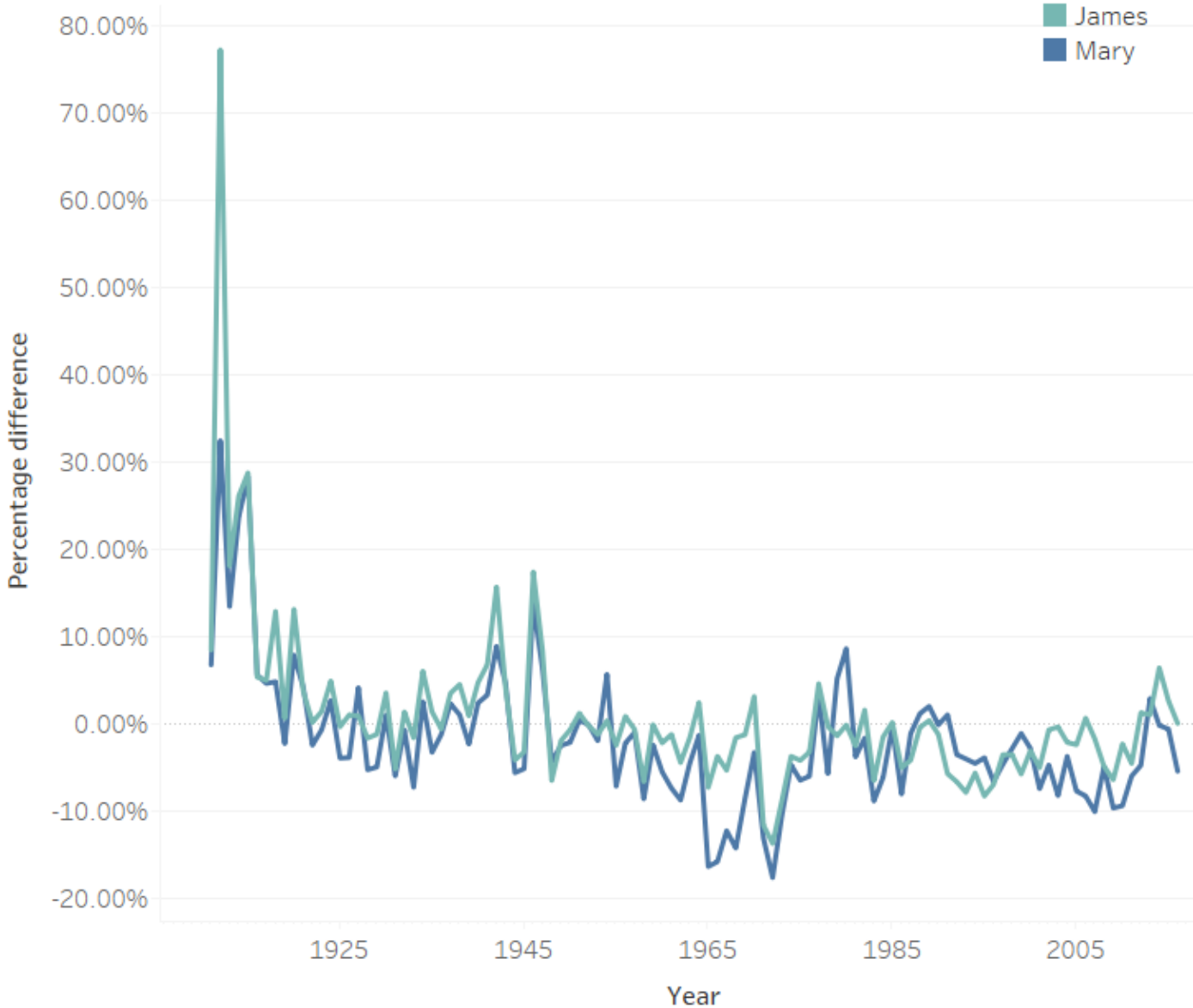


Percentage difference of name by year

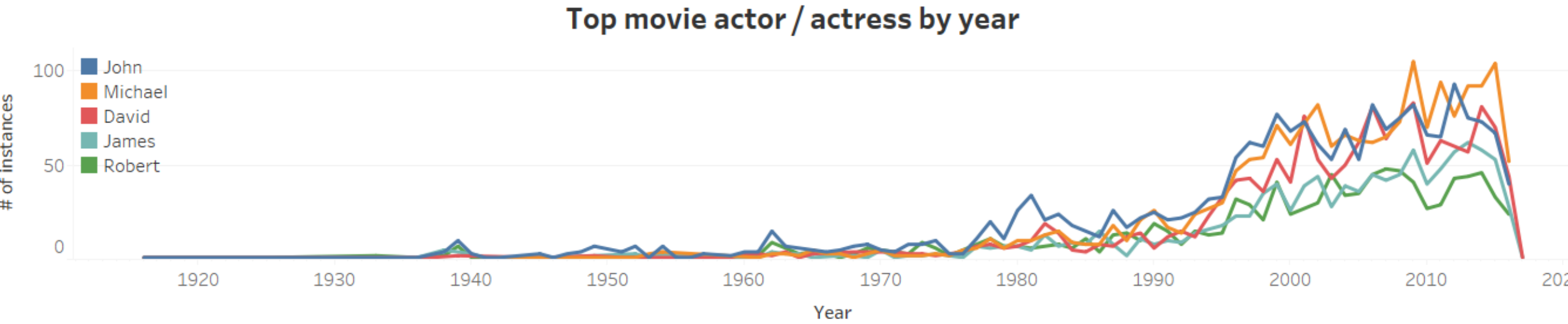
Top names by sex since 1910 - 2016

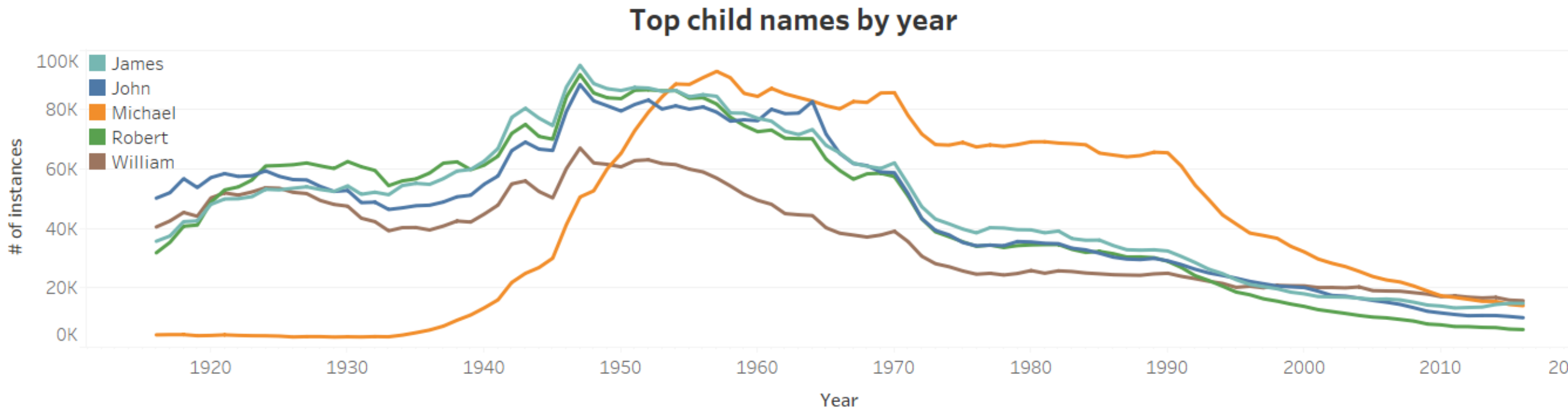


Mary & James percentage difference by year



Comparision with movie dataset





Weekly update Tableau dashboard approaches

1. **When data is on the client side:**
   - Use online sync client updates option which refreshes the data
   - The Tableau online workbook needs a live connection to the published data source
   - Need to manually click 'refresh' on the dashboard after the sync

2. **When data is published to Tableau Online:**
   - Extract Refreshes, Subscriptions, and Data-Driven Alerts
   - Enable Extract Refresh Scheduling and Failure Notification
   - Use "**tabcmd**" for weekly updates

3. **Use Tableau bridge:**
   - Maintain live connections to on-premises relational data
   - Schedule refreshes for extracts for On-premises data, Oracle, web data connector
   - For some web data connector data sources, scheduling refreshes is not an option due to OAuth

Critical points of failure

- Change in the data format (data types, shuffled attributes, etc.)
- Size of the data goes beyond the limit
- Latest data not reflected in all the worksheets
- When Tableau server cannot complete a scheduled refresh for five consecutive times, then it suspends the refresh
- Extract fails when login credentials does not get embedded while publishing the workbook

Resource requirements

- Minimum Hardware Requirement (8 GB RAM, Free Disk Space of 15GB)
- Database (SQL / NoSQL / Files / etc. based on the requirement)
- Tableau Bridge
- List of questions the dashboard should able to answer
- How frequently the data will update
- Data Source

Recommendation to drive sales for any Toy Company

- Identify top k states that have maximum number of children’s in 2016 (Example: California, Texas & New York are top 3 states have maximum children's)
- These 3 states have almost the equal number of boys and girls. So, a Toy company can promote all gender toys.
- If the company have the names of the children's via 3rd party / Cookie data, than the company can predict the gender using this data and can show them relevant toy Ad

Approach 2

- The trends in the birth rate can help the compnay to manufacture their toys in the right quantity.
- The company can also get rough estimation of child population growth in the coming years
- Introduced popular name based themes by giving discounts on toys(Example: Children's name with Mia in the California state will get extra 10% discount). We can do this for all states to drive more sales.

QUESTIONS?