

Challenge 4: Baby Names

By: Data In Motion, LLC

Scenario

You are a **data analyst** and you have been tasked with summarizing data about the occurrence of baby names. Your manager would like the following questions answered.

Get the Data

Download

dataset :<https://github.com/adishkhatri/Data-In-Motion-LLC-Data-Analysis-Challenges/tree/main/4.%20Baby%20Names>

1. How many different names exist in the dataset?

1.02M

Total Different Names

2. What is the name with most occurrences?

Jacob 242.87K

3. What is the standard deviation of count of names?

97.4

Standard deviation of Count of Name

4. Get a summary of the dataset with the mean, min, max, std and quartiles.

```
df.describe()
```

	Year	Count
count	1.016395e+06	1.016395e+06
mean	2.009053e+03	3.485012e+01
std	3.138293e+00	9.739735e+01
min	2.004000e+03	5.000000e+00
25%	2.006000e+03	7.000000e+00
50%	2.009000e+03	1.100000e+01
75%	2.012000e+03	2.600000e+01
max	2.014000e+03	4.167000e+03

```
df.quantile()
```

Year	2009.0
Count	11.0

- ☐ AK
- ☐ AL
- ☐ AR
- ☐ AZ
- ☐ CA
- ☐ CO
- ☐ CT
- ☐ DC
- ☐ DE
- ☐ FL
- ☐ GA
- ☐ HI
- ☐ IA
- ☐ ID
- ☐ IL
- ☐ IN
- ☐ KS
- ☐ KY
- ☐ LA
- ☐ MA
- ☐ MD
- ☐ ME
- ☐ MI
- ☐ MN

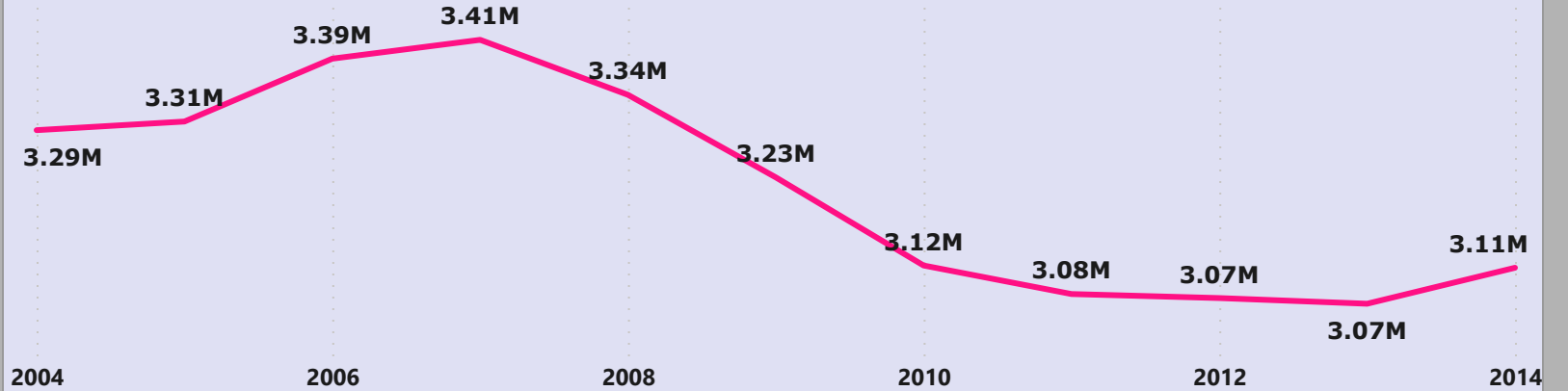
Dashboard

2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

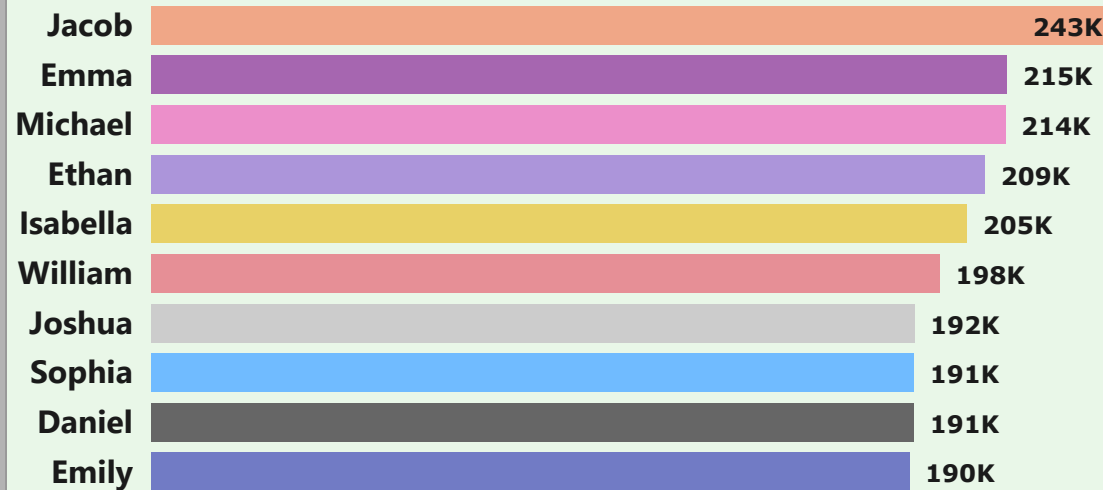
1.02M

Total Names

Name Used by Year



Top 10 Most Used Name



Count of Name by Gender

