

Data Dames

Paige Patrick, Dana Vera, Leigh Tanji, Yejin Hwang, Mathena Abramson

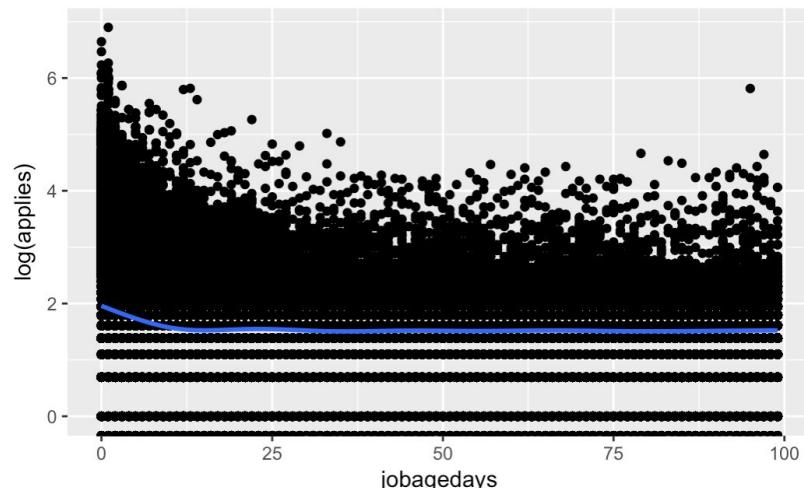
How to increase the number of applications for a specific job post?

To model or not to model?

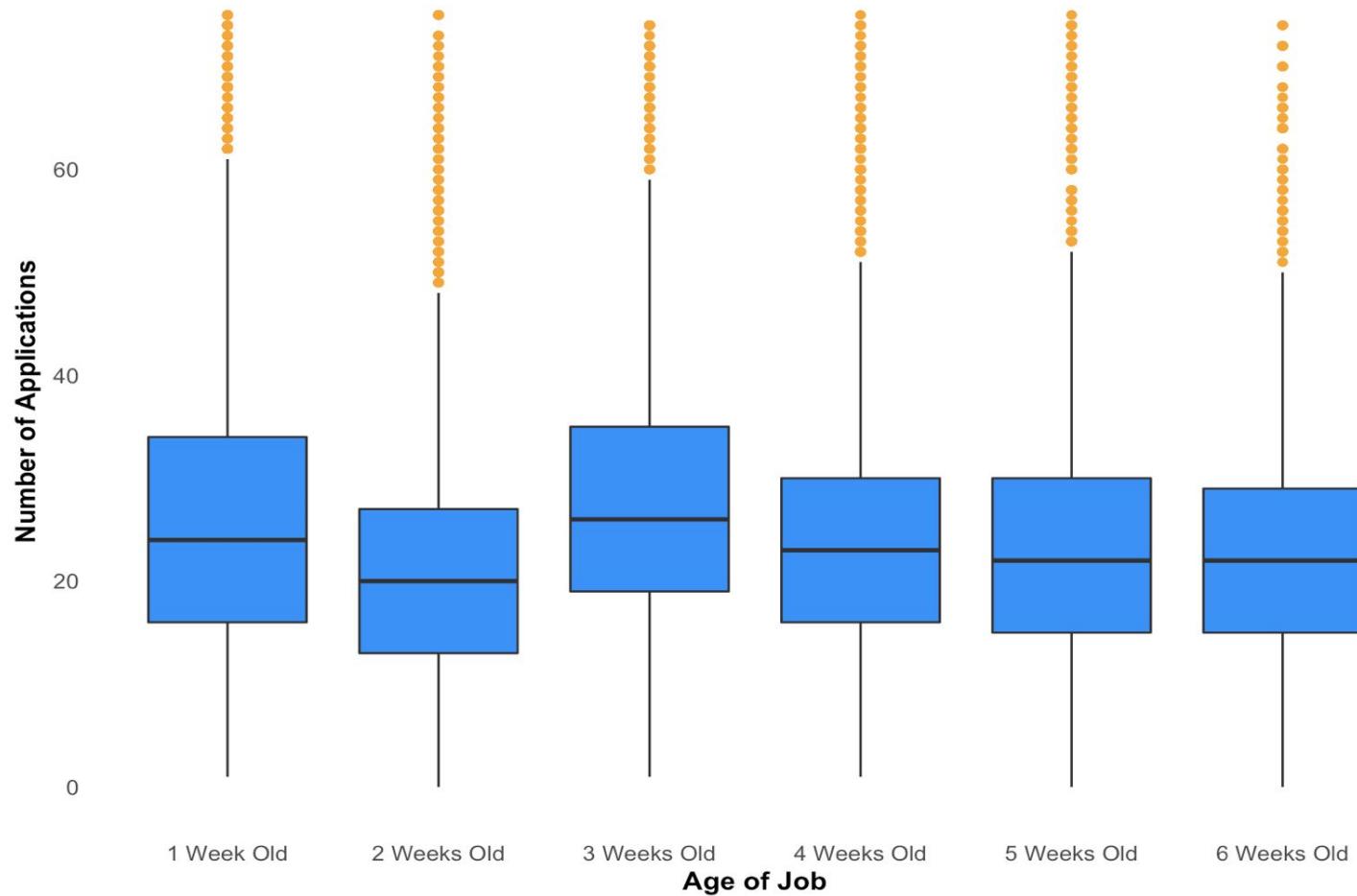
- > Simple Linear Regression
- > Multiple Linear Regression
- Transformations
- > Mixed Models

When does a job go stale?

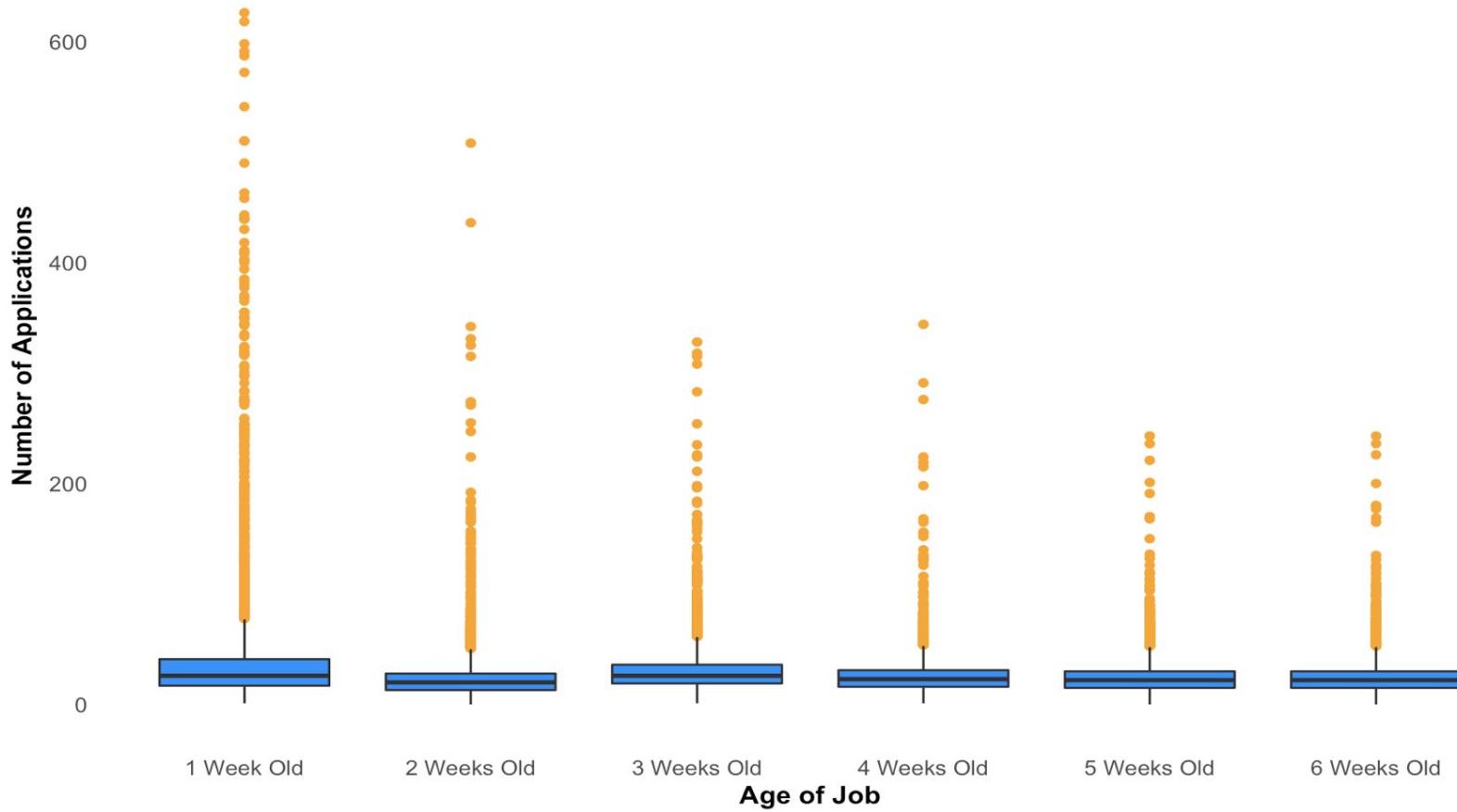
> Relationship between Number of Applications with Time



Number of Applications Per Week



Number of Applications Per Week



What's going on?

Confounding Variables

- > Profile companies
- > Sponsorship
- > New Users to Indeed

Outliers

- > Most likely sponsored jobs
- > Application rate is decreasing across weeks
- > Need more variables

The screenshot shows the Indeed job search interface. The search bar at the top has 'Recent Graduate' in the 'What' field and 'CT' in the 'Where' field. Below the search bar, it says 'Job Title, Keywords Or Company' and 'City, State, Or Zip'. To the right are buttons for 'Find Jobs' and 'Advanced Job Search'. The main content area displays search filters on the left: 'Recent Graduate jobs in Connecticut', 'Recommended Jobs - 63 new', 'My Recent Searches' (listing 'research scientist - Connecticut', 'staffing firm - Connecticut - 2 new', and 'Recent Graduate - 156 new'), '» clear searches', 'SORT BY: Relevance - Date', and 'Salary Estimate' (\$70,000). The results list shows two job posts: 'Scout for Admissions - New' from 'Mode Enterprises, Inc - Rocky Hill, CT 06067' (marked as 'Sponsored' and 'day ago') and 'CLINICAL NURSE - New' from 'Company with Recent Graduate jobs' (posted '1 day ago'). A red circle highlights the 'Sponsored' status of the first job post. On the right side of the results, there is a sidebar with a message 'Be the first to see new Recent Graduate jobs in CT', an input field for 'My email:', a checkbox for 'Also get an email with jobs recommended just for me', and a blue 'Activate' button.



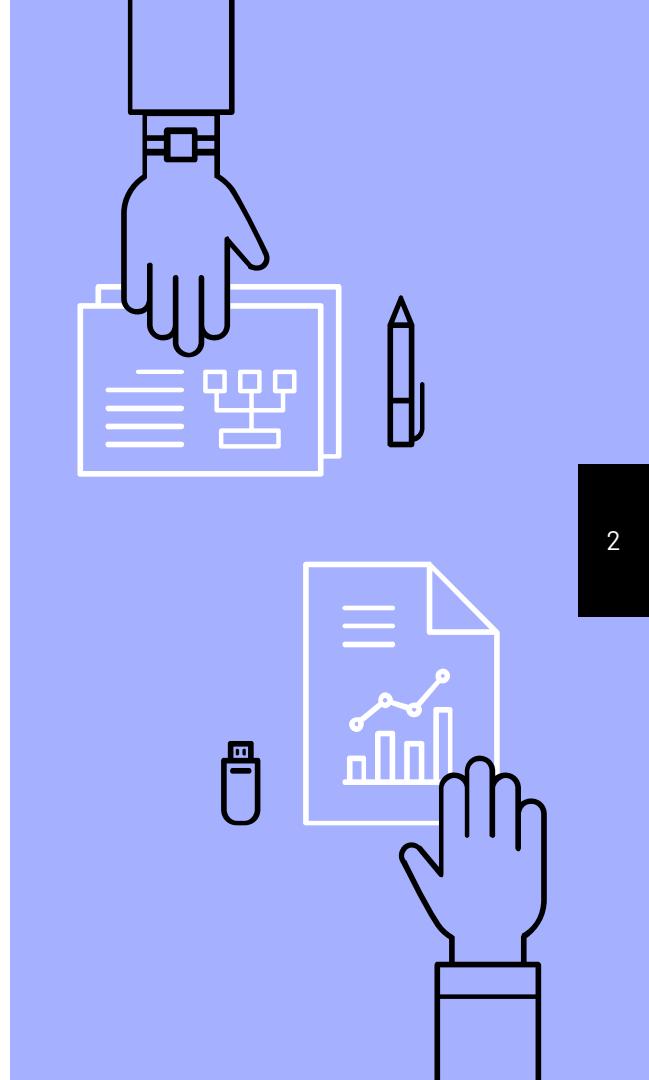
POPPFest 2018

ACOF

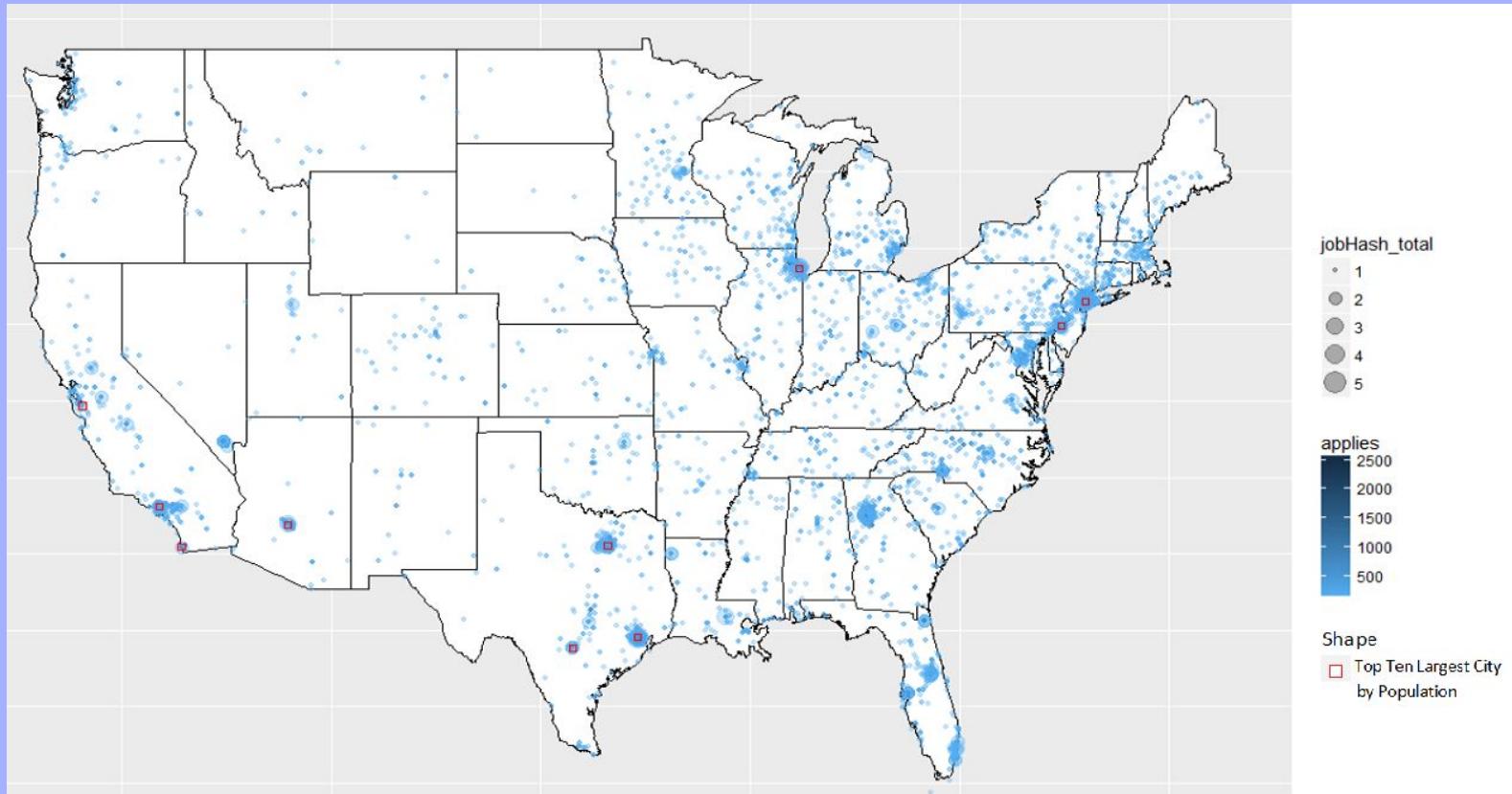
Natalia Khoudian, Bodhi Nguyen,
Brendan Seto, Jason Seto, Kelly Yang

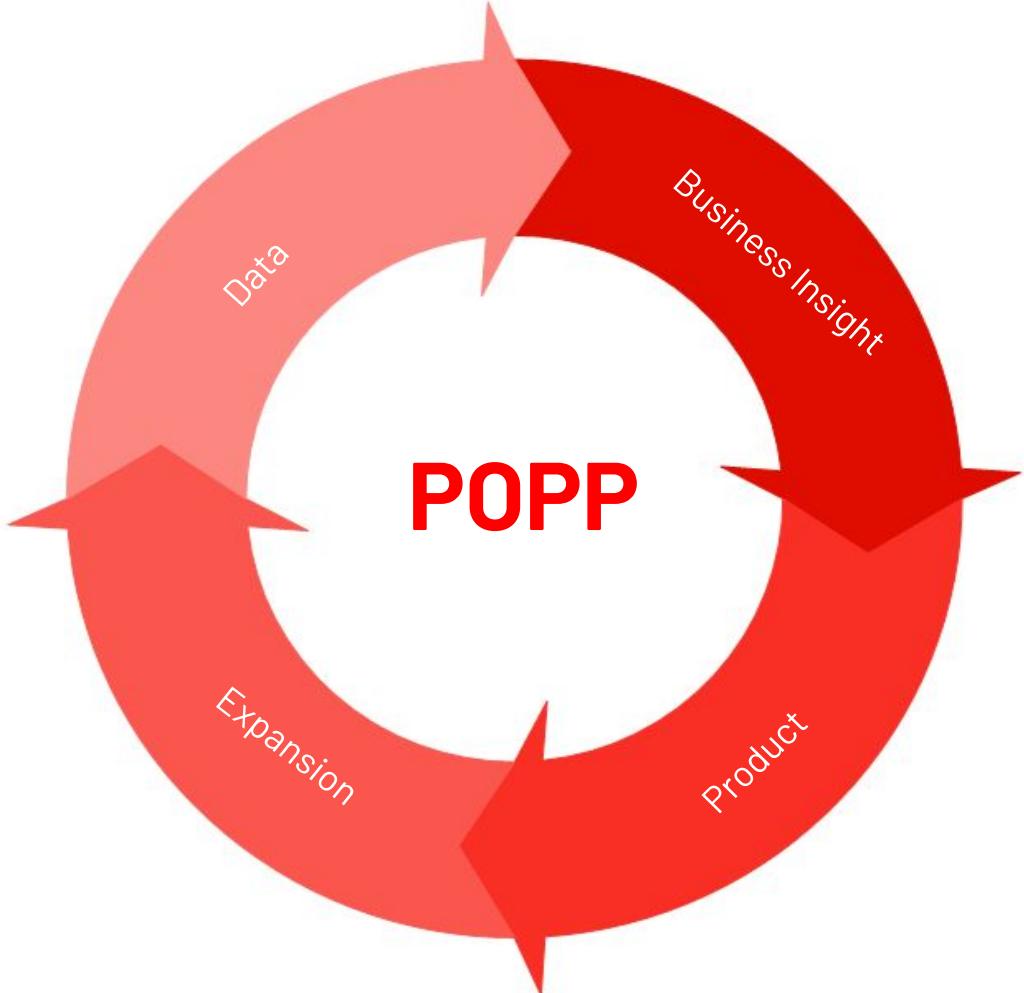
Personalized Occupation Posting Program (**POPP**)

- ▶ Predictive model lacked power
- ▶ Also too impersonal
- ▶ How can we help indeed.com help employers (their clients)?
- ▶ **POPP!** - an application to offer personalized posting insights



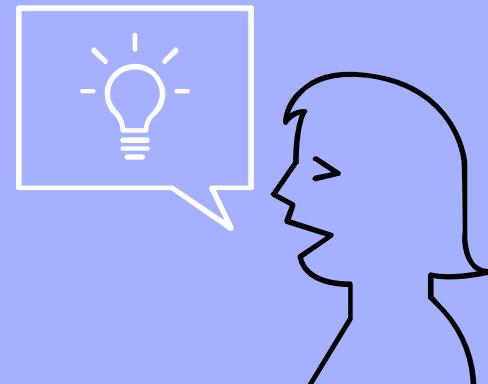
Map of Jobs Applications





NEXT STEPS

- ▷ Gather data to increase the possibilities of POPP
- ▷ Enable businesses to maximize clicks and applies per position
- ▷ Streamline the job posting process

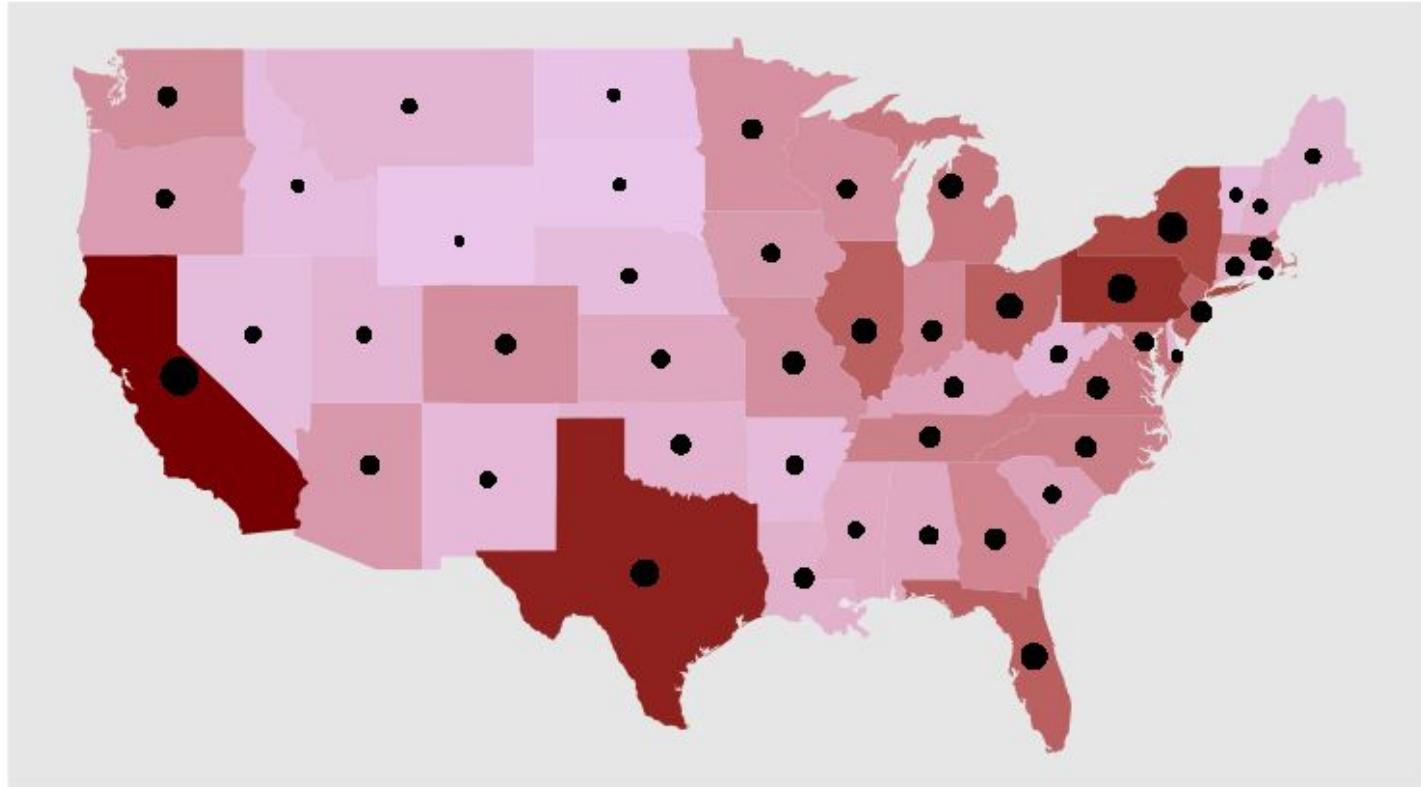


Career Insights for High School Students

R2-D2, Smith College

Yuqing Geng, Nichole Yao, Lizette Carpenter, Emily Kim

Geographical Distribution of Institutions and Jobs



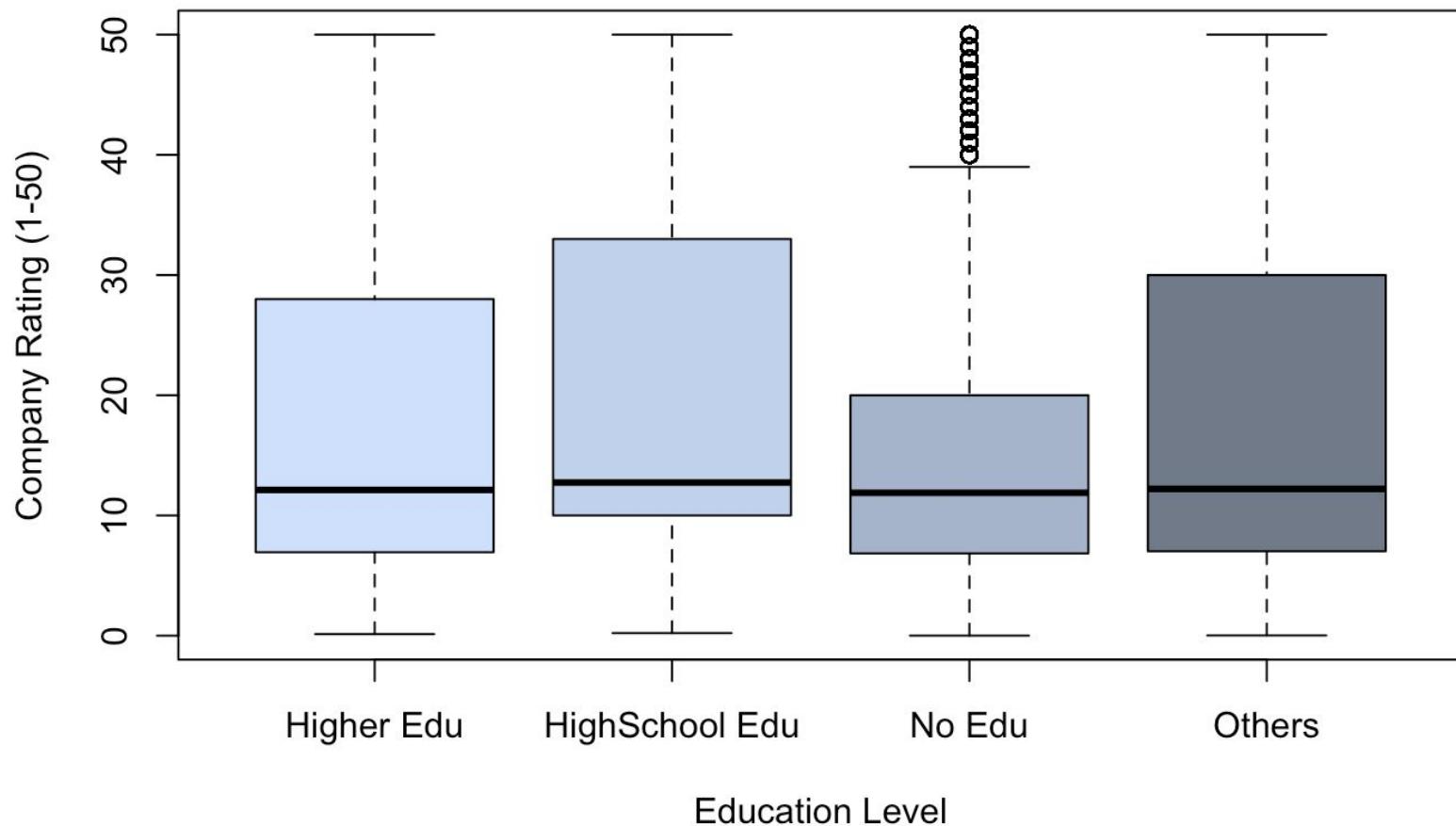
Number of Jobs (in millions)

20
40
60
80

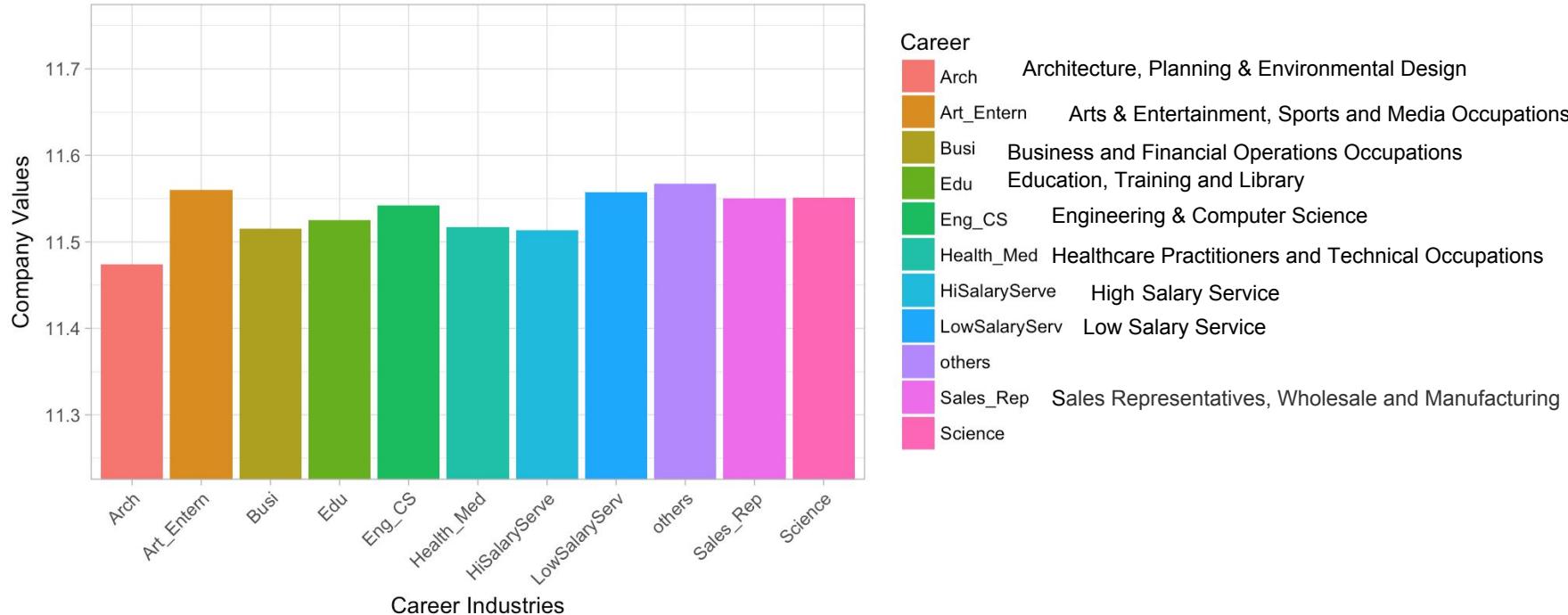
Number of Institutions

● 250
● 500
● 750
● 1000

Required Education Level and Company Value



Company Values in Different Career Industries



User Application for Hiring Season Investigation



The Code Girls

Rachel Bostick

Sumatra Dhimoyee

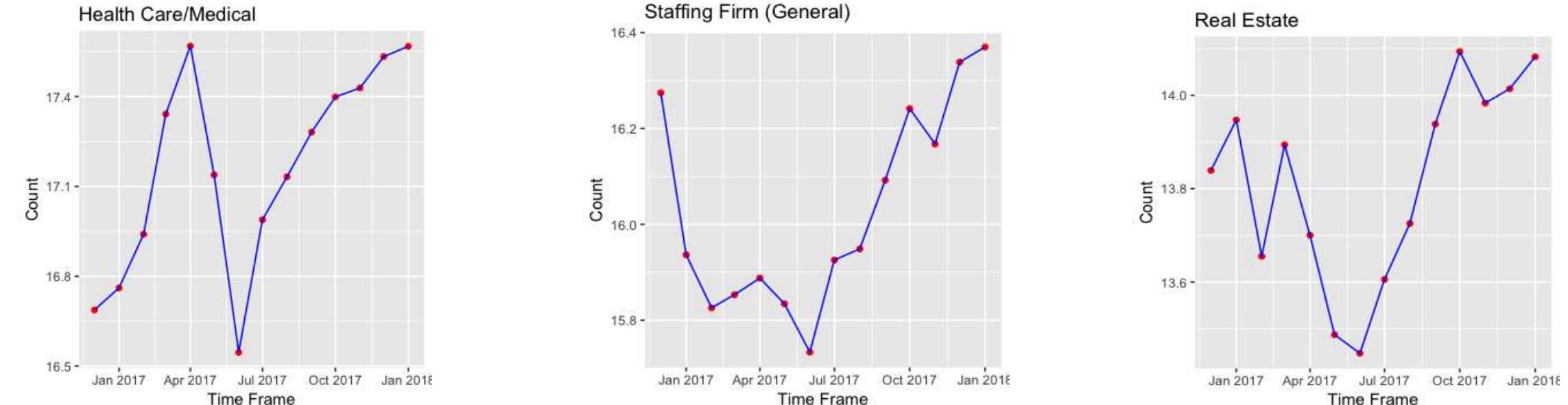
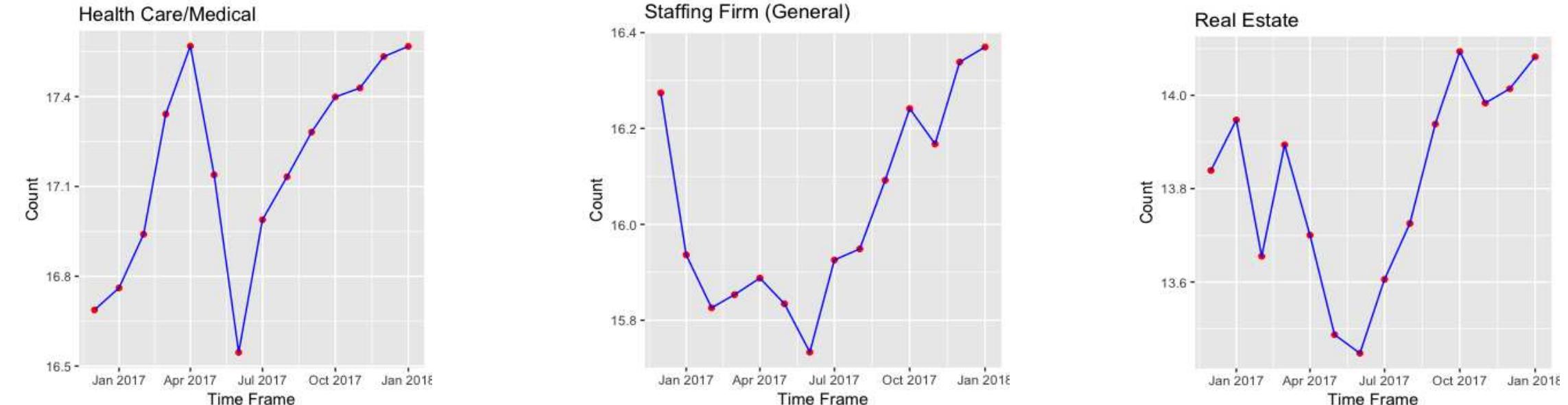
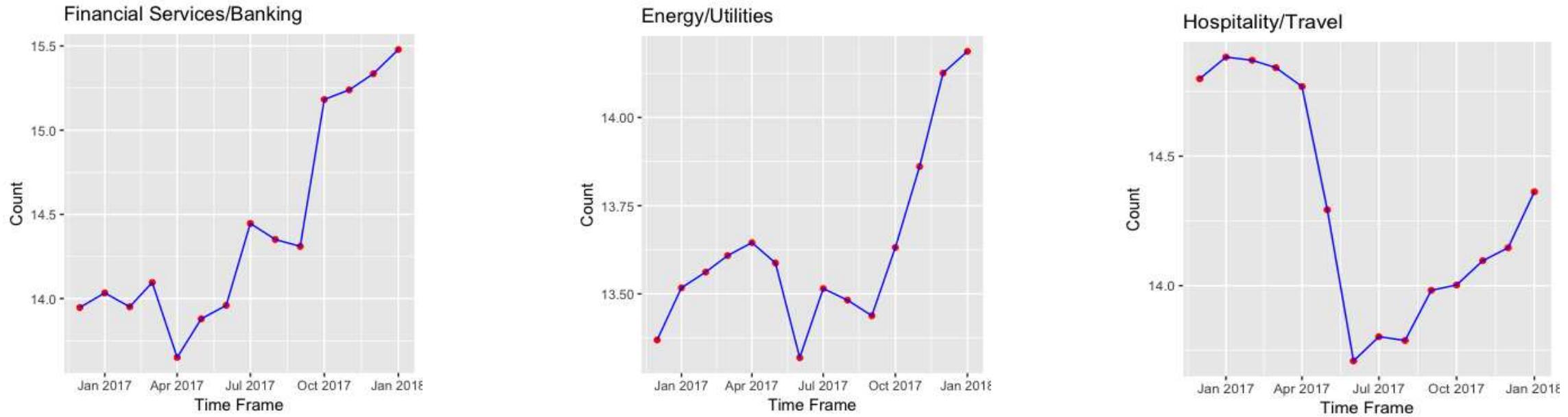
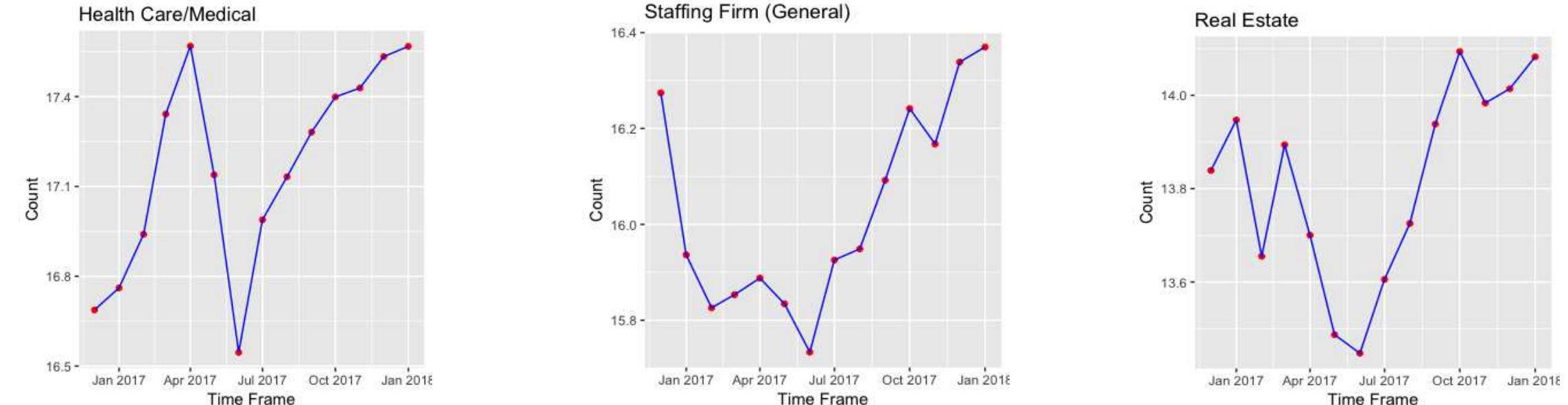
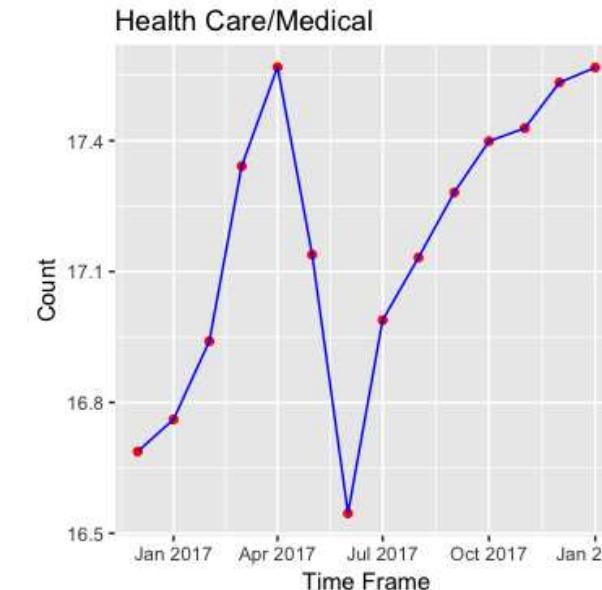
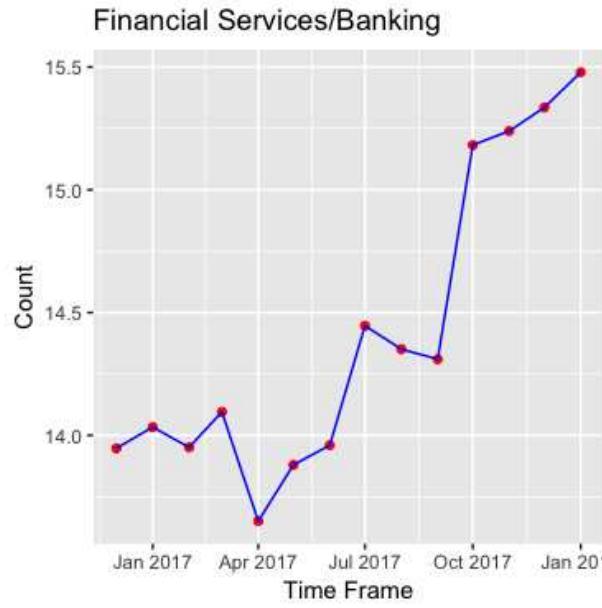
Amelia Johnson

Anisa Kabir

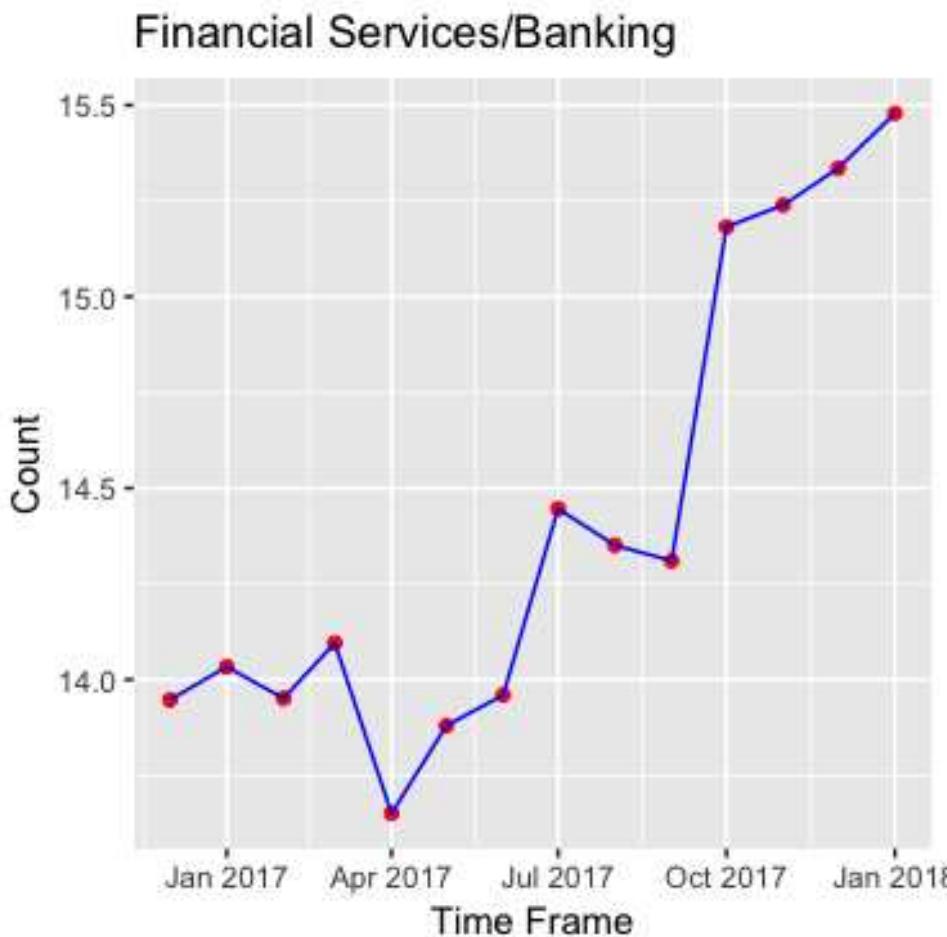
Raeesa Mehjabeen

Industries with overall positive job posting growth trend

$$y = \log\left(\sum_{i=1}^n employerJobCount\right)$$



The Financial and Banking Industry



- Percentage increase in Job Posting count and seasonal trend observed for highest clicked from September 2017 – October 2017

139%

Confounding Variables

- Changes in Indeed usage per industry
- Industry growth or decline
- Overall economy unemployment decrease in 2017

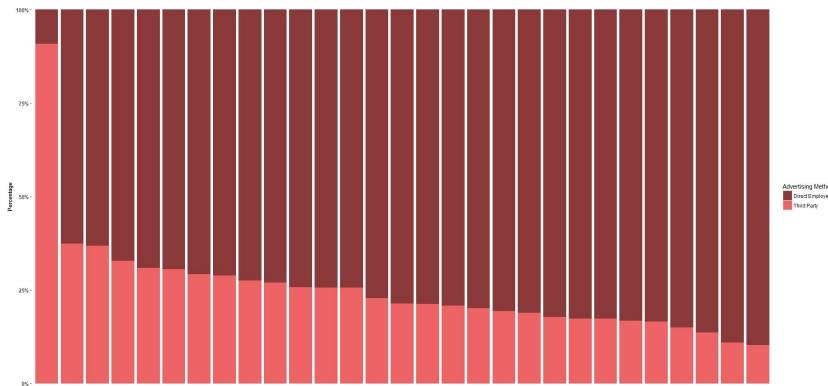
Conclusion:

We believe our app can be applied to the Indeed website to provide a simple interface for users to track hiring seasons and hotspots for their industry of interest. For further development:

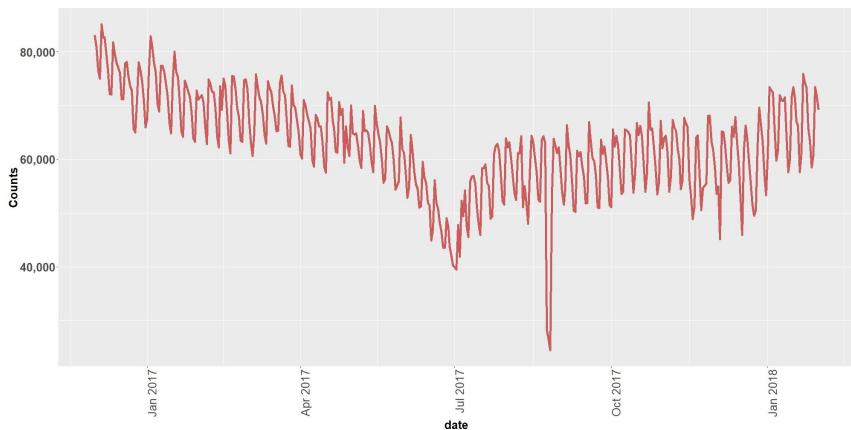
- Interactive maps
- Larger data would fine tune our time-series graphs
- Live data for time-series graphs

DataFest 2018

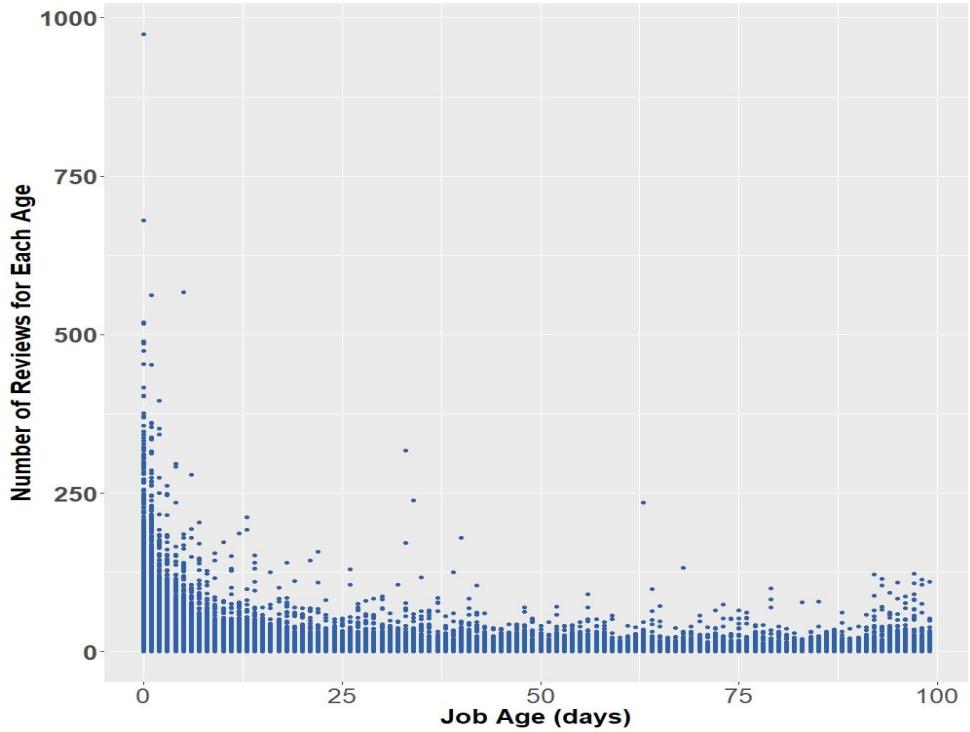
The **indeed**® User: Employers and Job Seekers



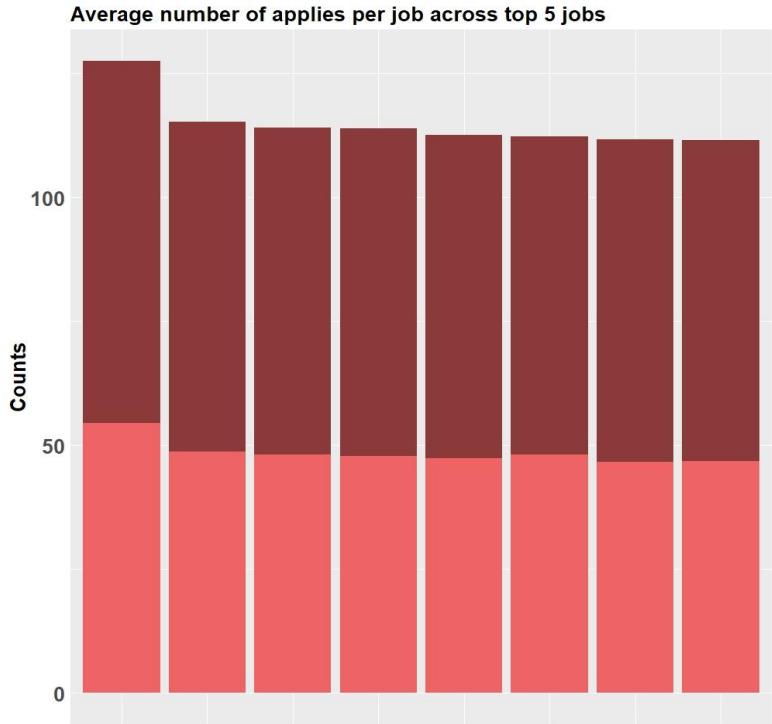
Most job postings are **small firms** (< 1000 employees), especially **0-49 employees** and are **direct employers**



The recruitment season is **late fall to early spring**

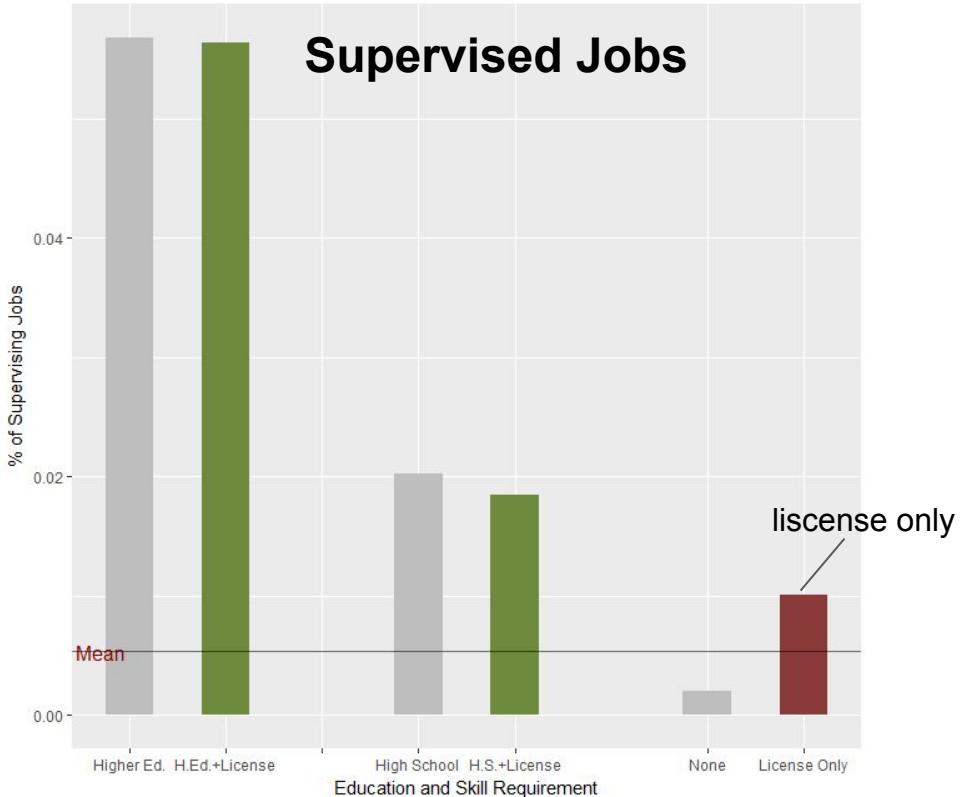
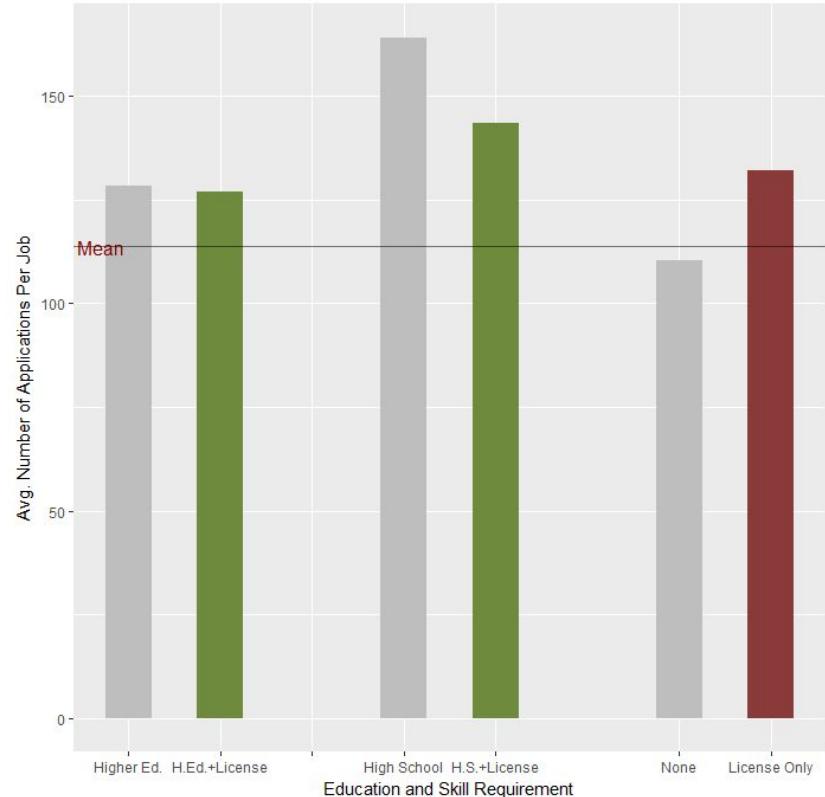


Most app reviews are done during the first 15 days of the lifetime of the jobs → applications submitted **within the first 15 days has higher chance of getting reviewed** → **APPLY EARLY!**



Number of apps submitted and apps reviewed across all job categories is relatively consistent = ~115 job app/job opening and ~40% get reviewed → same for most industries

Further Insight into Indeed Job Seekers - When in Doubt, Get a License!!!



```
Call:  
lm(formula = avg_app ~ licenseRequiredJob + edutype + jobcategory +  
    advtype, data = supply_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.006037	-0.003270	-0.001615	0.001562	0.283652

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.102e-03	2.064e-04	24.721	< 2e-16 ***
licenseRequiredJob	-1.025e-03	8.983e-05	-11.412	< 2e-16 ***
edutypehighschool	-5.277e-05	7.957e-05	-0.663	0.50720
edutypenone	-1.446e-04	5.657e-05	-2.556	0.01060 *
jobcategoryClerical support workers	6.109e-04	2.100e-04	2.910	0.00362 **
jobcategoryElementary occupation	1.104e-04	2.037e-04	0.542	0.58771
jobcategoryManagers	1.024e-04	2.000e-04	0.512	0.60881
jobcategoryProfessional	-3.830e-05	1.999e-04	-0.192	0.84804
jobcategoryService and sales workers	1.887e-05	2.007e-04	0.094	0.92510
jobcategorySkilled Workers	-8.251e-05	2.018e-04	-0.409	0.68261
jobcategoryTechnicians	2.523e-05	2.048e-04	0.123	0.90196
advtypeJob Board	-5.352e-04	7.077e-04	-0.756	0.44949
advtypenone	4.688e-04	2.761e-05	16.977	< 2e-16 ***
advtypeOther	-1.410e-03	1.568e-04	-8.991	< 2e-16 ***
advtypeStaffing Agency	-1.181e-03	2.810e-05	-42.044	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.005434 on 232032 degrees of freedom

(3981 observations deleted due to missingness)

Multiple R-squared: 0.01317, Adjusted R-squared: 0.01311

F-statistic: 221.2 on 14 and 232032 DF, p-value: < 2.2e-16

- License significant
- Advtype significant
- Education requirements offset by other factors

Recommendations for Indeed, Employers, and Job Seekers

This is

indeed

ASA

DATA

657373656e7469616c6c792c206
16c6c206d6f64656c7320617265
2077726f6e672c2062757420736
f6d65206172652075736566756c

FIVE COLLEGE

657373656e7469616c6c792c206
16c6c206d6f64656c7320617265
2077726f6e672c2062757420736
f6d65206172652075736566756c

FEST

2018

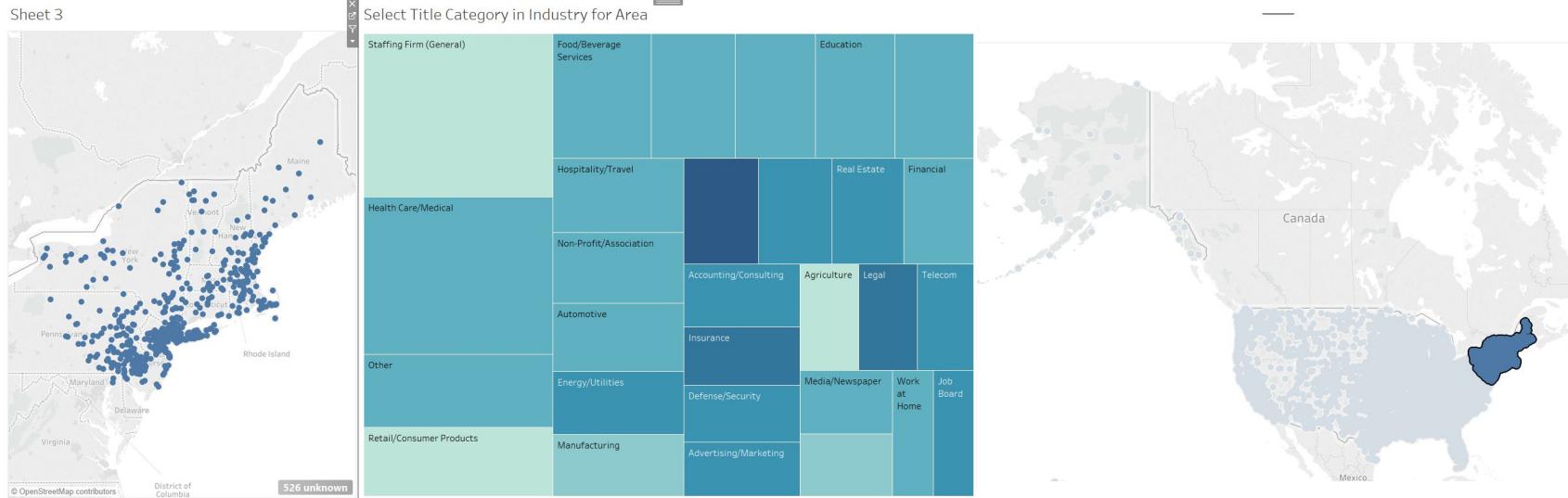
the

presentation by

Jared Nussbaum, Katrina Greene, David Welch, and Robin Wu

Employees

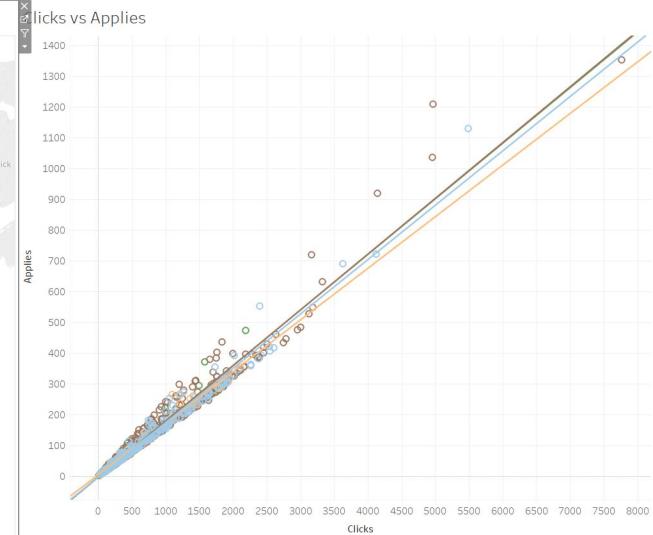
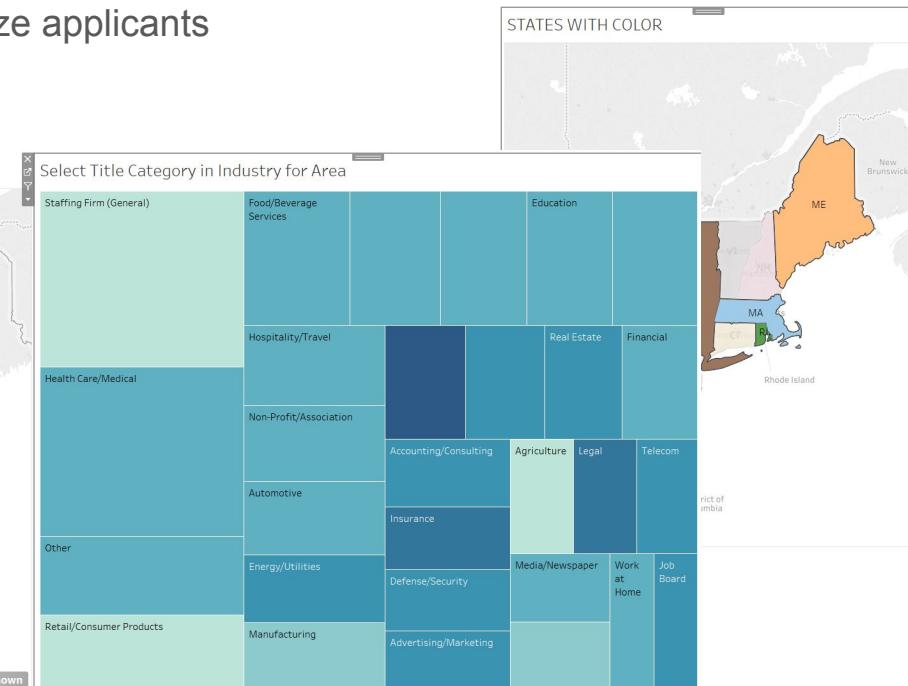
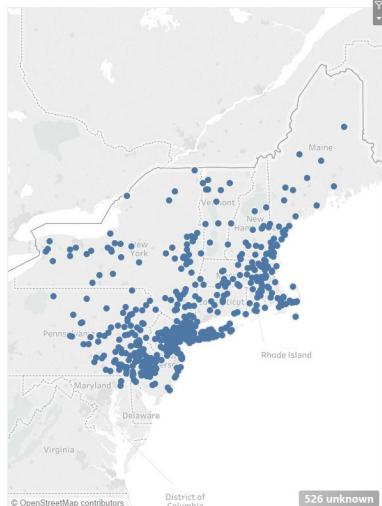
- Prospective employees searching for jobs via indeed.com
- How do aspiring workers maximize their ability to get a job through Indeed?
 - Help them find the jobs they like by...
 - Organizing jobs visually



Companies

- Companies posting jobs and looking for employees on indeed.com
- How do companies maximize their ability to hire employees via Indeed?
 - Maximize applicants

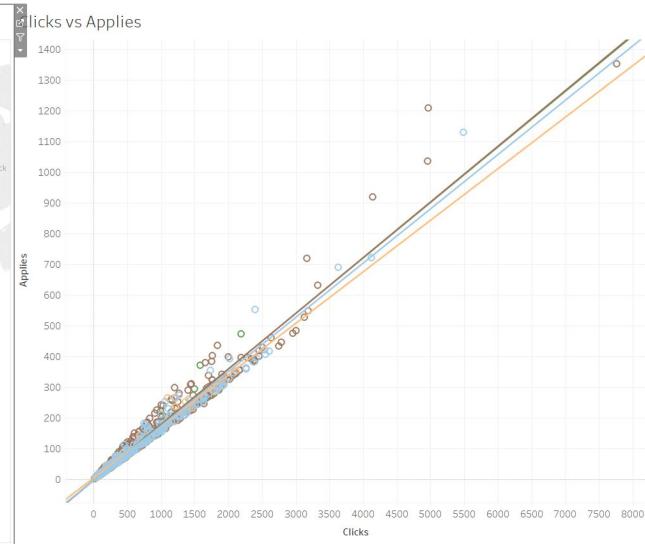
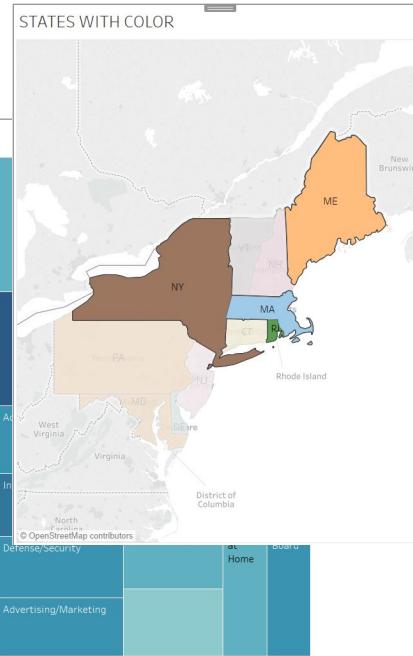
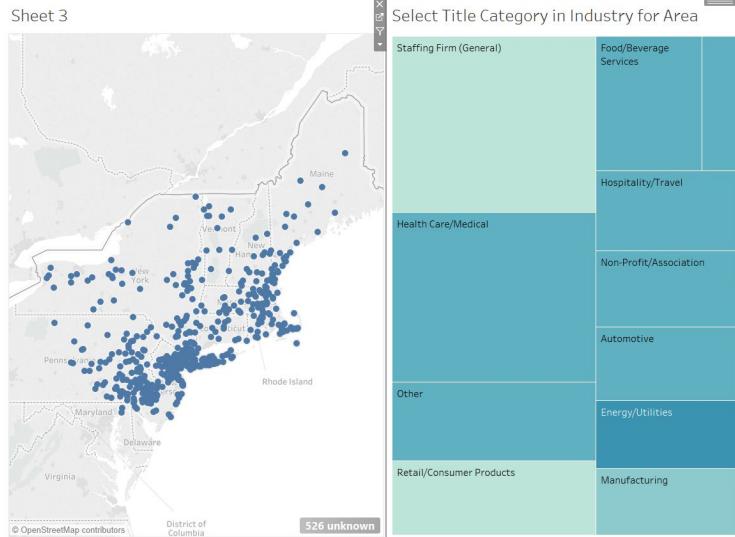
Sheet 3



Indeed

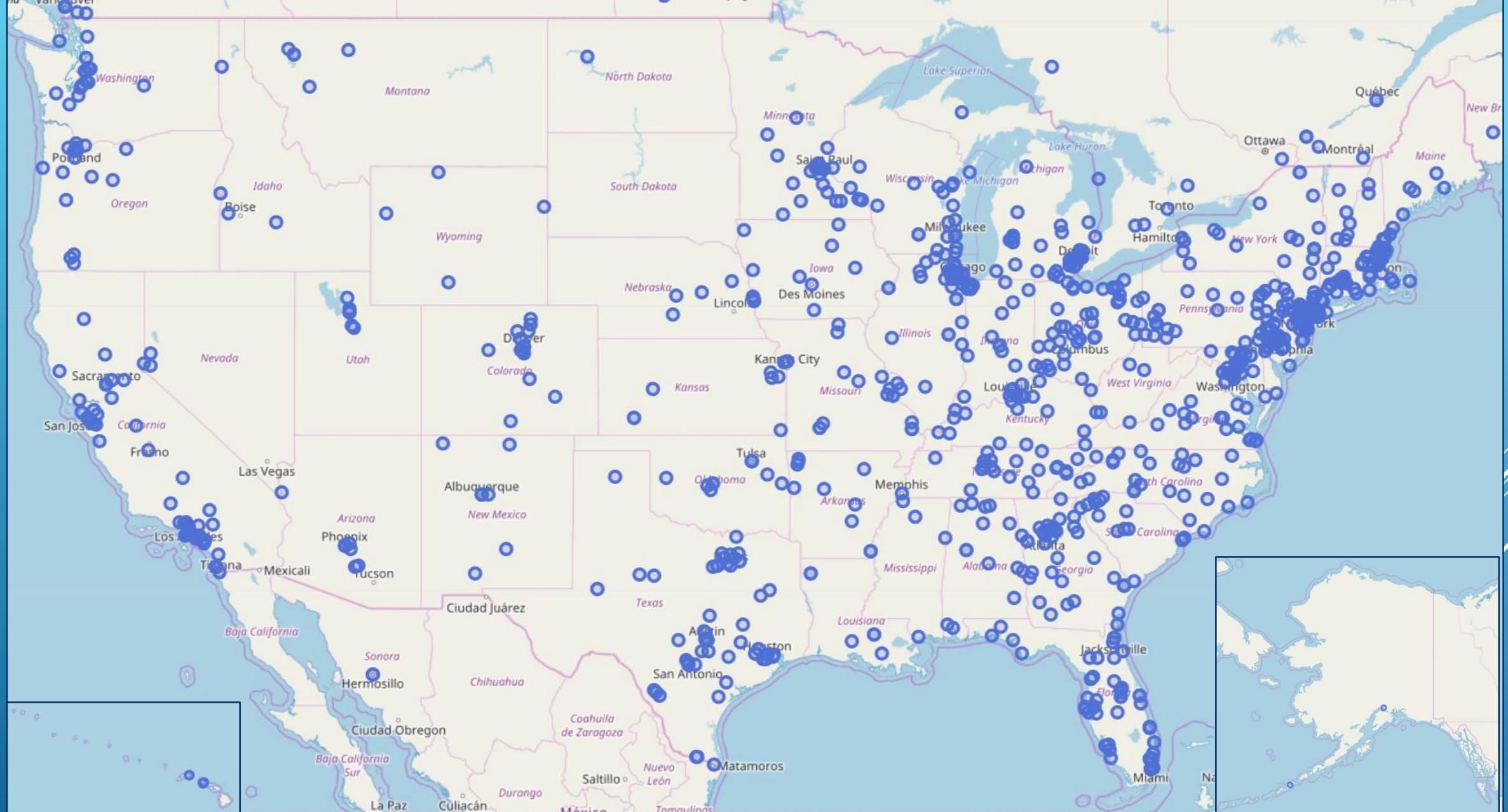
- Indeed.com themselves
- How can we get as many people as possible to use indeed?
 - Make it successful
 - Market like heck

Sheet 3

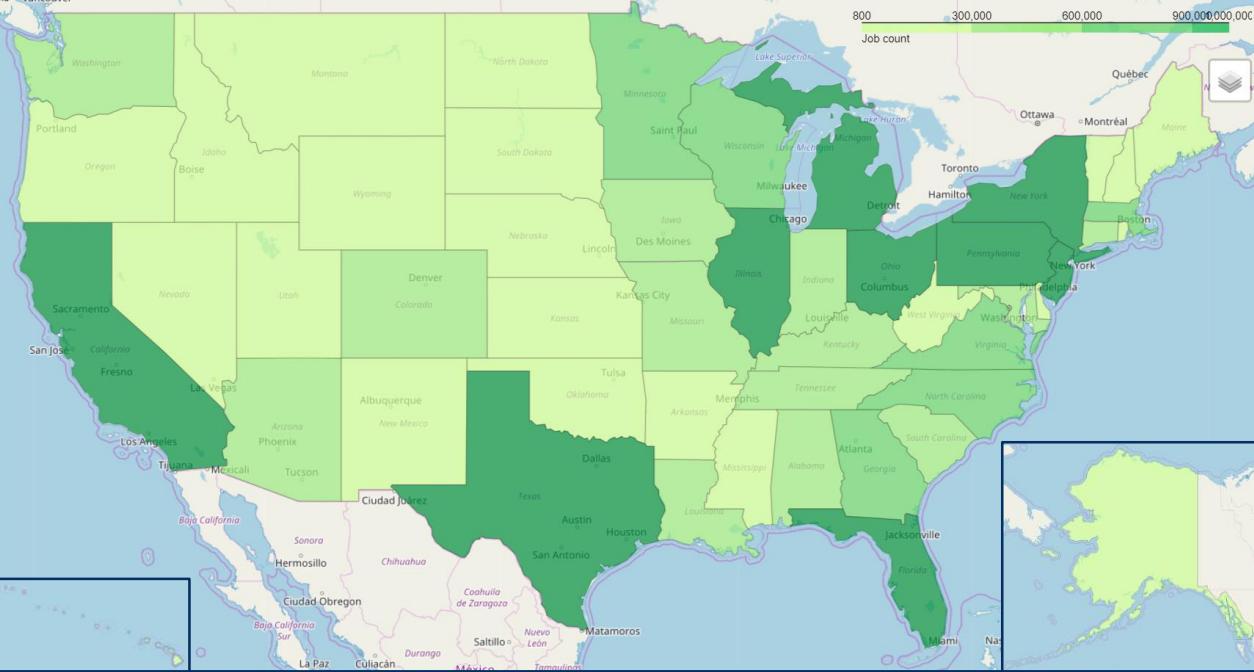


TEAM 23333

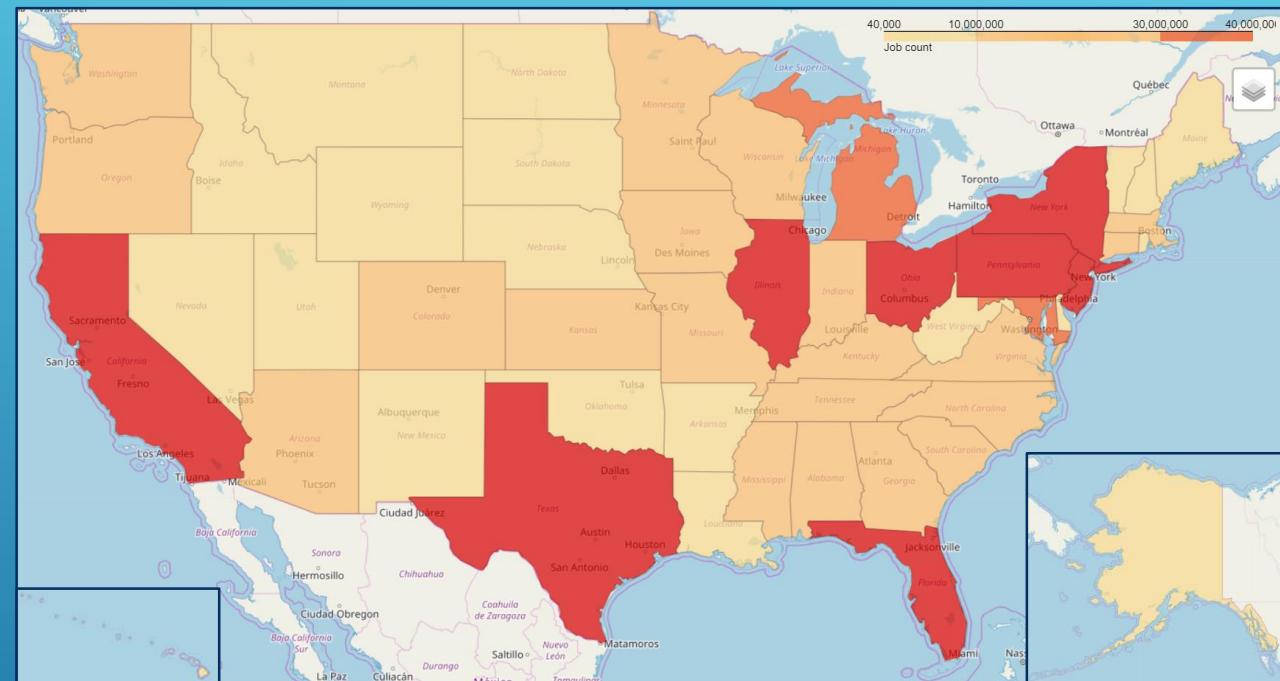
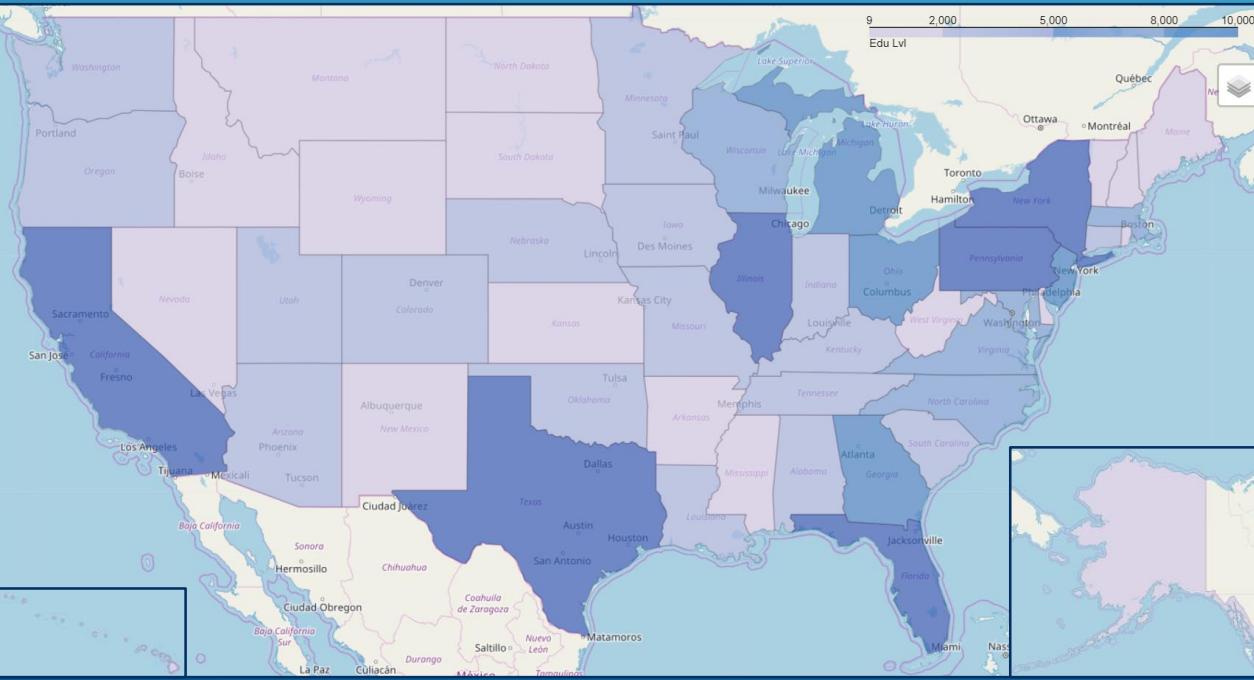
Team members: Ziwei He, Hao Gao,
Huan Wang, Fusheng Yang



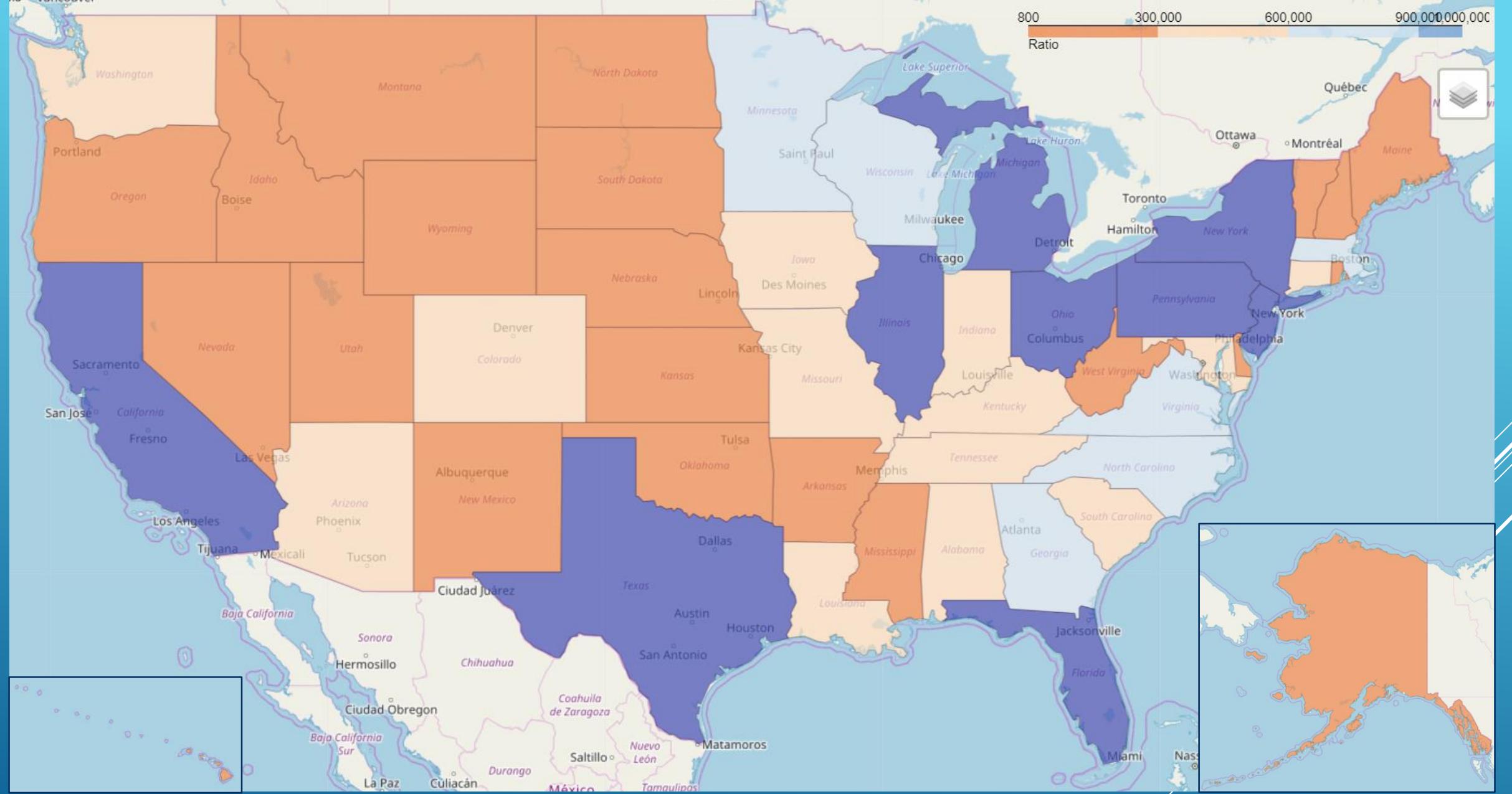
Company Location



Applies ↑
Education Requirement ↓



Job Amount ↑



Job Amount vs. Applicant Amount

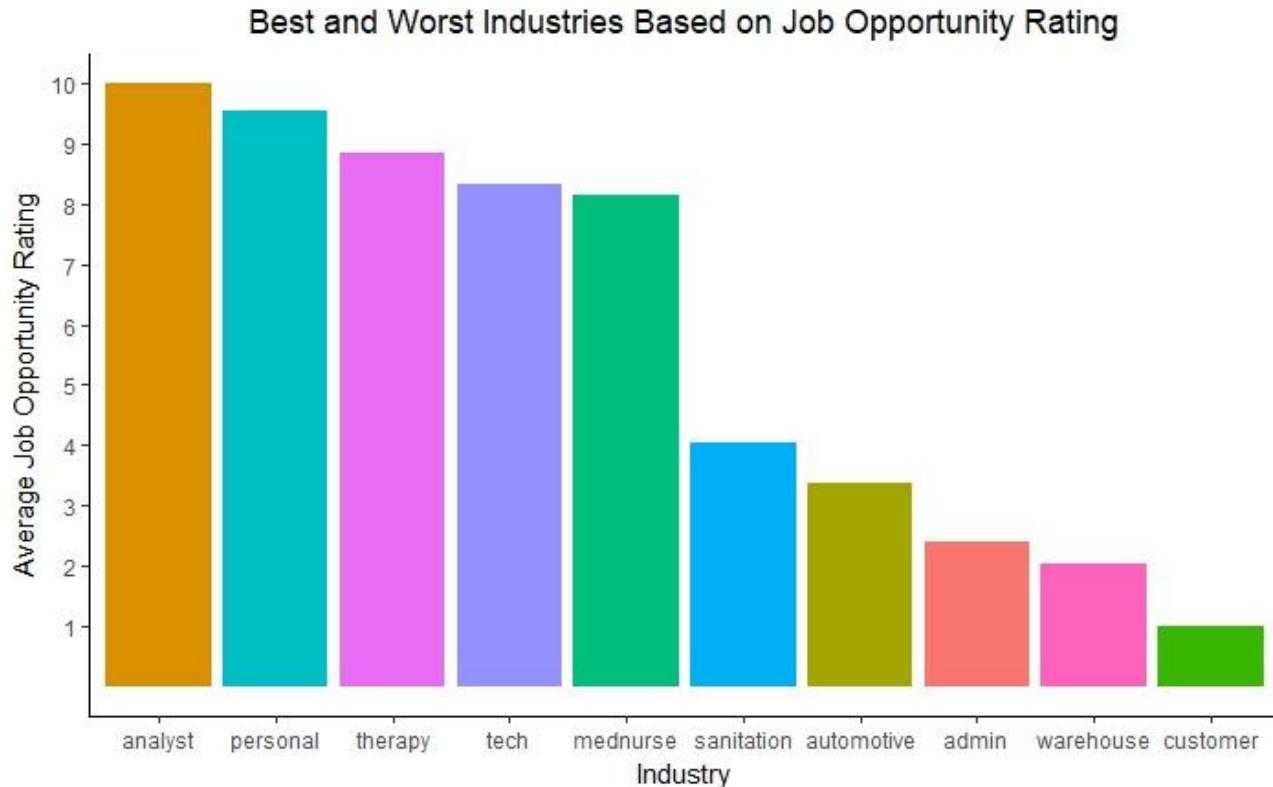
Indeed: Clicks, Applies, and Job Opportunity

Regression Toward the Team

Munkh-Erdene Baatarsuren, Ryan Cox, Evan Moore, Ilina Shah, Kristina Yamkovoy

Methods

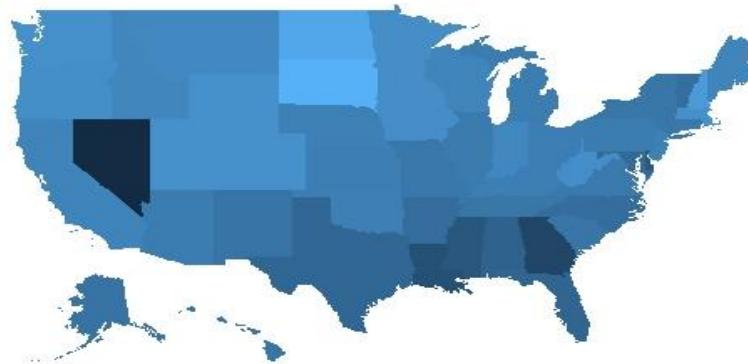
- Clicks → applications
→ vacant roles are filled.
- *Job opportunity rating* describes percentage of unique postings in industry divided by percentage of average daily clicks for that industry, normalized on scale of 1-10.



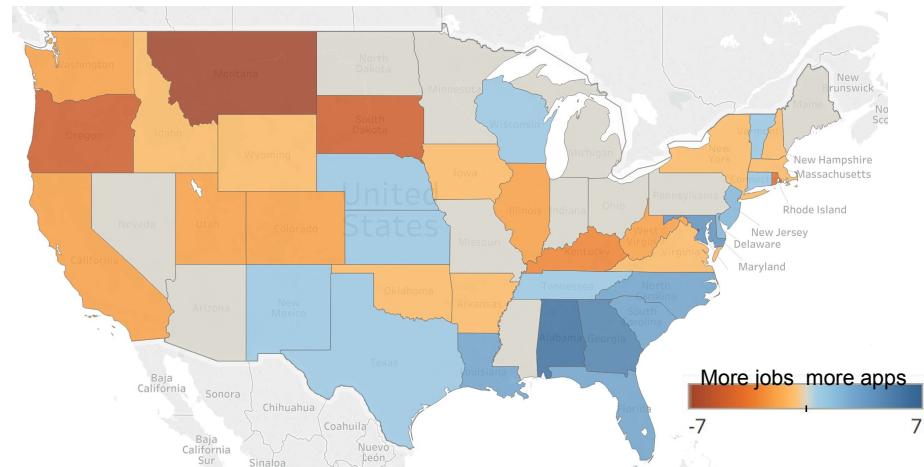
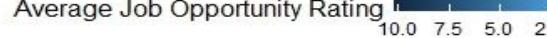
State Level Results

- First plot focuses on higher education jobs. Second plot focuses on license and no requirement jobs.
- *Employees*: Focus on dark areas indicating high potential for employment.
- *Employers*: Focus on light areas indicating high potential for growth.
- **Conclusion:** Nevada/Georgia good for higher ed requirement, Montana/Oregon good for no requirement.

Average Job Rating By State (Top 5 Jobs only)



Average Job Opportunity Rating

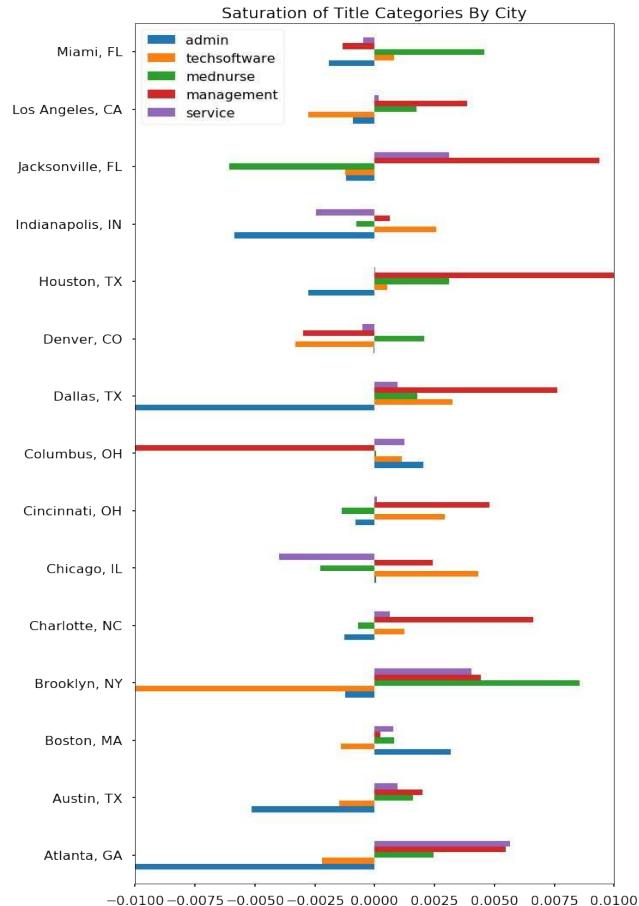


More jobs = more apps



City-Level Results

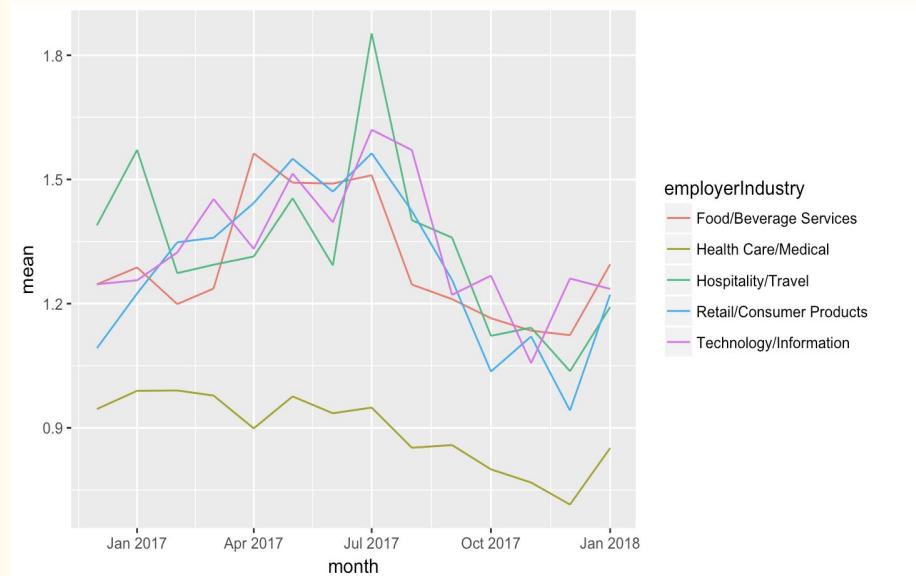
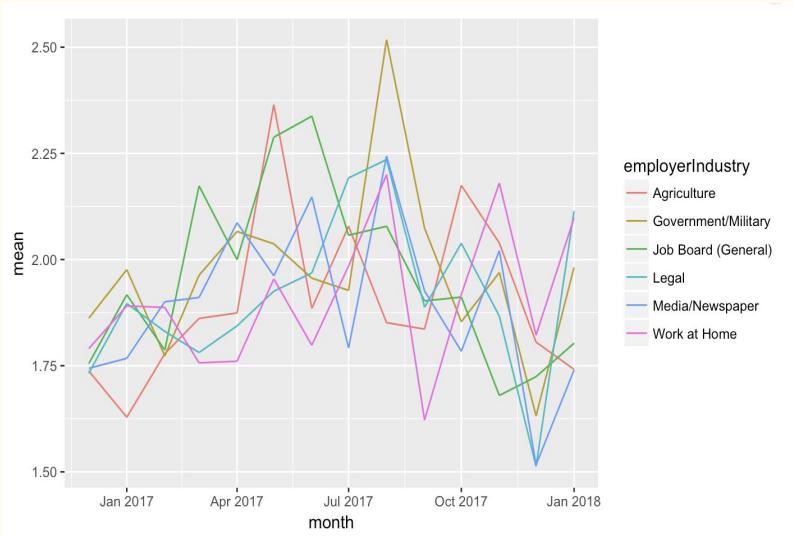
- Determined a “saturation level” based on a difference of normalized “click interest” and “job availability”
- X axis: proportional difference
- Y axis: city
- **Conclusion:** Companies seeking to expand ops might want to look at cities where the desired workforce is undersaturated



Employment through Indeed and Visa Obtainment

Raeesa Alam, Meredith Pan, Caroline Li, Qiaqia Ji, Naila Arskey

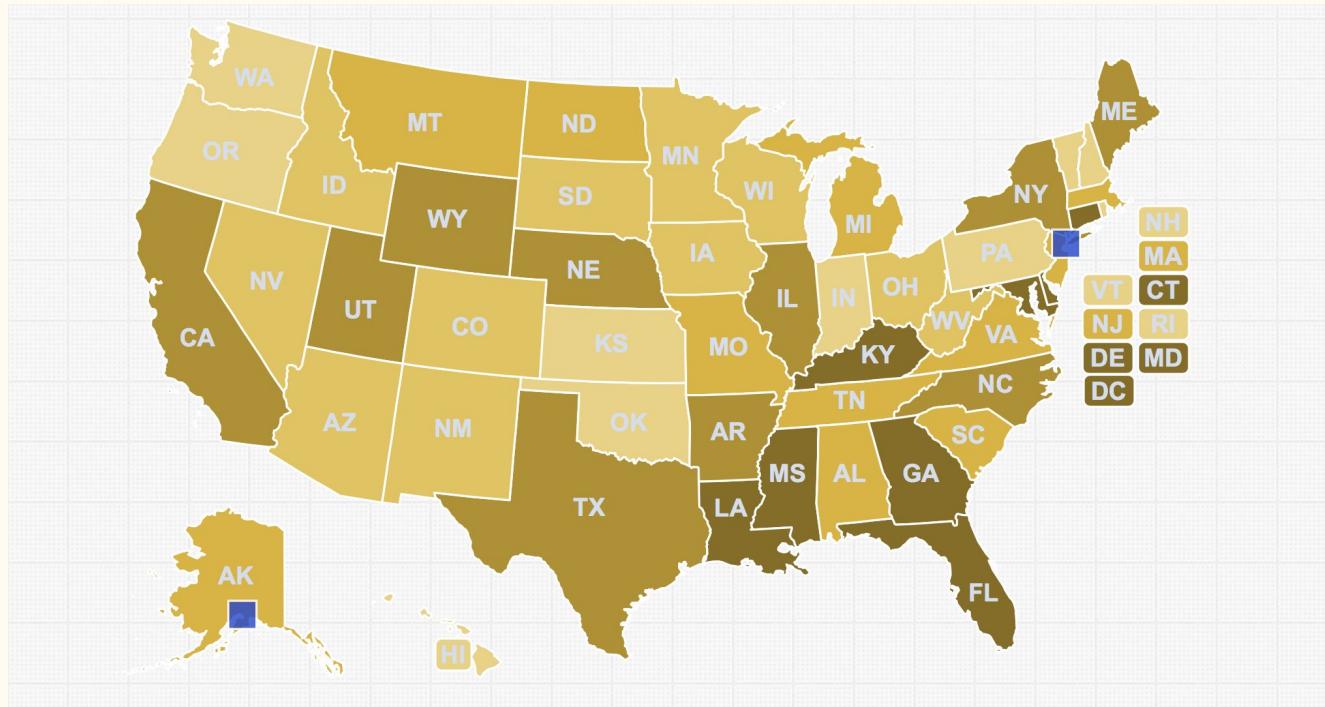
Which Industries have the most Competition?



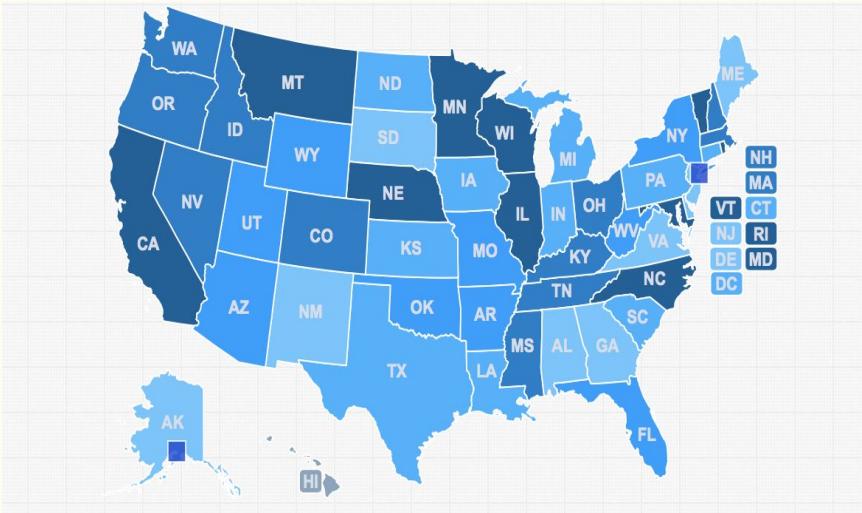
Ratio of Job demand to States

0.007 - 1.79(GA)

Possible outlier:
3.040(NP)

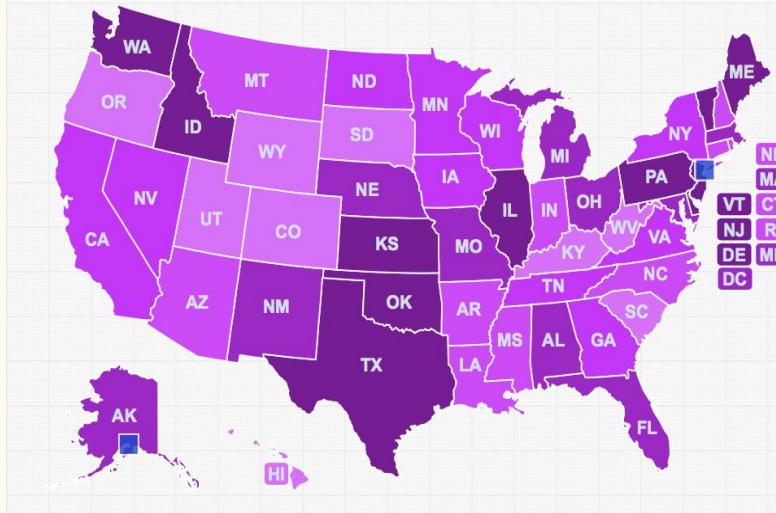


Where are most profitable
jobs for H-1B holders?



72,214(VI, WY) - 86,907(VT, PR)

Where will an international
student be most likely to
obtain a H-1B?



70.97%(AK, WY) - 96.77%(PR, VT)

How affective is the review system?

5C DATAFEST 2018

Datachamp 2

**Kendra Clemenzi
Priscilla Lu
Tim Raxworthy
Tony Zhang**

Let's look at how it works...

The company review you see

★★★★★ 20,606 reviews -

Each employee gives an individual rating

Employees give a rating on each component

Job Work/Life Balance

Compensation/Benefits

Job Security/Advancement

Management

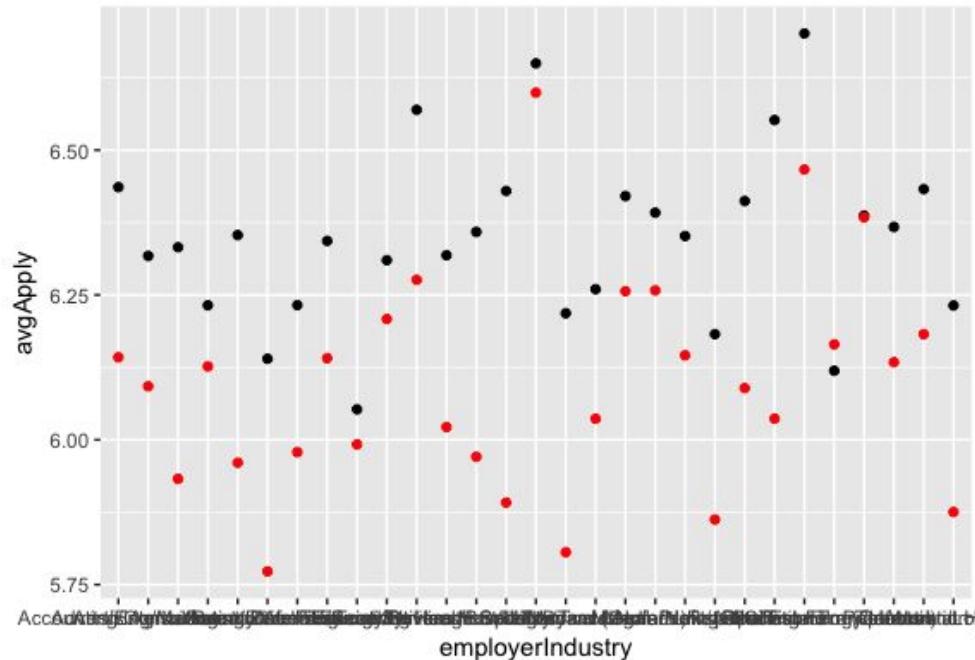
Job Culture

Company with rating
vs
Company without rating

1,274,120 VS 3,004,791

30%

Is it helping the employees????



Black dot:
company
without rating

Red dot:
company with
rating

What can we change?

Pros

Cons

*MAKE KEYWORD HASHTAGS!

*Make rating for each component for obvious and accessible.

Preprocessing Data

- Selected a number of random samples from the entire dataset
 - Reservoir sampling to uniformly and randomly select samples
 - Used sample sizes of 10,000 and 500,000 samples
- Select a number of columns used as features to be used for machine learning models
 - Choose data that are related by job details and popularity
 - Average overall company rating, Description Length Chars, Clicks, Applies, Education Requirements, Cities, States
- Longitude and Latitude references for Cities
 - Use Google Maps to read in longitude and latitude strings
 - Show education requirement trends in United States regions

Machine Learning Models

- Decision Trees
 - Non ensemble learner for random forest classification. Mainly splits off one feature of the data. The split for children nodes is not always binary(20% accuracy)
- Random Forest Classification
 - Ensemble learning method consisting of many classification trees on subsets of the entire data set(10% accuracy)
- Support Vector Machines(Classification)
 - Minimize the distance between each data point with the line. Only able to run for the 10,000 samples(10% accuracy)

Data Visualization

- We tried to make a data visualization through querying google maps API
- By querying google maps API, we are able to get the longitude and latitude points of each major city
- Preprocessing the data through grabbing certain columns with querying proved to be too much
 - Columns queried: noEducationRequirementsJob, highSchoolEducationRequirementsJob, higherEducationRequirementsJob

For the Future

- Try grabbing more data and features to be able to distinguish between different classes
- The accuracy of the model increased with more data but the processing time also increased
- Use cluster to speed up computation to balance workload versus wait time
- Supply company activity information to users
- Analyze job type and date posted in relation to current events