

CS 329E Final Project

Natasha Halfin, Evan Johnston, Kaylyn Zvonek

Industries' Relationship to Employment and Population

We were interested in looking at factors that have relationship to a state's population and employment. We compared data from different industries against state employment and housing data.

Data Used

- Top 5000 U.S. Companies
 - Industry, growth, and revenue
- State employment data
 - Number of firms and establishments, and payroll
- State housing data
 - Population, number of houses, land and water area, population density

Data Used

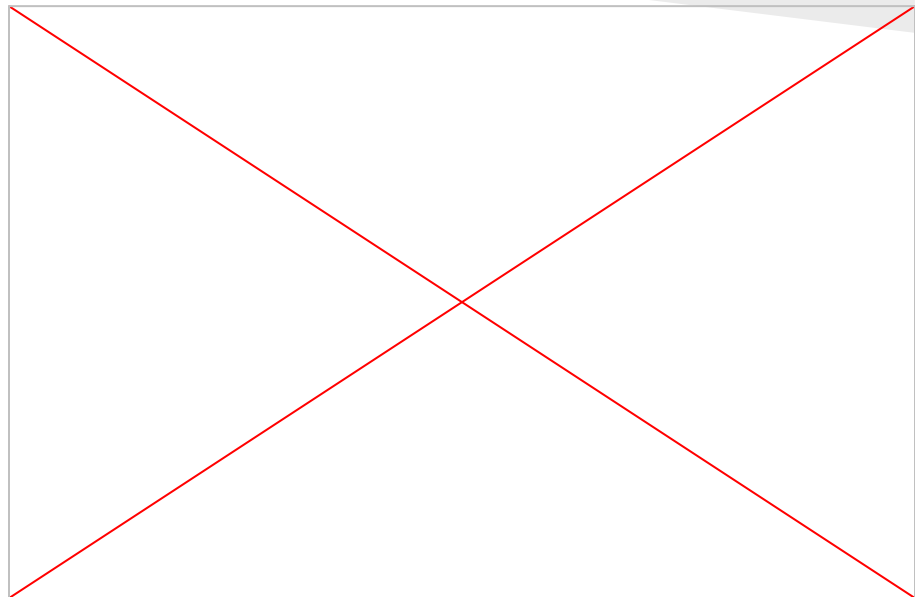
Top Companies

##	COMPANYID	STATE	INDUSTRY	REVENUE	GROWTH
## 1	173	NJ	Health	5400000	11.100
## 2	174	PA	Health	9500000	11.098
## 3	175	AZ	Logistics & Transportation	25400000	11.048
## 4	176	VA	Government Services	39000000	11.029
## 5	177	MA	Insurance	207700000	10.999
## 6	178	UT	Software	15500000	10.971

Employment

##	STATE	NUM_FIRMS	NUM_ESTABLISHMENTS	EMPLOYMENT	PAYROLL
## 1	AL	73661	97938	1585761	61284931
## 2	AK	16656	20427	258219	13986403
## 3	AZ	101205	131375	2134252	90838616
## 4	AR	50412	64815	978519	35633588
## 5	CA	701899	864913	12952818	700102379
## 6	CO	126478	153112	2035803	97691175

Housing



State Employment and Housing

- Left join of housing and employment data
- Mutated columns
 - Employment rate
 - Ratio of water to total area
 - Ratio of land to total area
 - Average number of employees per establishment

##	STATE	NUM_FIRMS	NUM_ESTABLISHMENTS	EMPLOYMENT	PAYROLL	POPULATION
## 1	AL	73661	97938	1585761	61284931	4779736
## 2	AK	16656	20427	258219	13986403	710231
## 3	AZ	101205	131375	2134252	90838616	6392017
## 4	AR	50412	64815	978519	35633588	2915918
## 5	CA	701899	864913	12952818	700102379	37253956
## 6	CO	126478	153112	2035803	97691175	5029196
##	HOUSING	TOTALAREA	WATERAREA	LANDAREA	POPULATIONDENSITY	HOUSINGDENSITY
## 1	2171853	52420.07	1774.74	50645.33	94.4	42.9
## 2	306967	665384.06	94743.10	570640.94	1.2	0.5
## 3	2844526	113990.30	396.22	113594.08	56.3	25.0
## 4	1316299	53178.55	1143.07	52035.48	56.0	25.3
## 5	13680081	163694.73	7915.52	155779.22	239.1	87.8
## 6	2212898	104093.67	451.78	103641.89	48.5	21.4
##	EMP_RATE	WATER_PERC	LAND_PERC	AVG_EMP	AVG_PAY	
## 1	0.3317675	0.033856116	0.9661438	16.19148	12.82182	
## 2	0.3635704	0.142388595	0.8576114	12.64106	19.69275	
## 3	0.3338934	0.003475910	0.9965241	16.24550	14.21126	

Top Companies

- Grouped by industry and state
- Summarized columns
 - Average revenue
 - Average growth

##	INDUSTRY	STATE	AVG_REVENUE	AVG_GROWTH
## 1	Advertising & Marketing	AL	3200000	0.422000
## 2	Advertising & Marketing	AR	16050000	2.785000
## 3	Advertising & Marketing	AZ	19558333	1.564167
## 4	Advertising & Marketing	CA	4119204933	3.886627
## 5	Advertising & Marketing	CO	13337500	2.338625
## 6	Advertising & Marketing	CT	9550000	3.001833

Top Companies in Texas

- Grouped by industry and state
- Summarized columns
- Filtered by State == TX

##	INDUSTRY	STATE	AVG_REVENUE	AVG_GROWTH
## 1	Advertising & Marketing	TX	17688462	6.641577
## 2	Business Products & Services	TX	27817143	2.497714
## 3	Computer Hardware	TX	19800000	1.294750
## 4	Construction	TX	38806452	2.373097
## 5	Consumer Products & Services	TX	24133333	1.518867
## 6	Education	TX	31700000	1.148200

Joining All Data

- Company data grouped by state
- Joined with housing and employment data
- Mutated columns from joined housing and employment data

##	STATE	NUM_FIRMS	NUM_ESTABLISHMENTS	EMPLOYMENT	PAYROLL	POPULATION	
## 1	AL	73661	97938	1585761	61284931	4779736	
## 2	AK	16656	20427	258219	13986403	710231	
## 3	AZ	101205	131375	2134252	90838616	6392017	
## 4	AR	50412	64815	978519	35633588	2915918	
## 5	CA	701899	864913	12952818	700102379	37253956	
## 6	CO	126478	153112	2035803	97691175	5029196	
##	HOUSING	TOTALAREA	WATERAREA	LANDAREA	POPULATIONDENSITY	HOUSINGDENSITY	
## 1	2171853	52420.07	1774.74	50645.33	94.4	42.9	
## 2	306967	665384.06	94743.10	570640.94	1.2	0.5	
## 3	2844526	113990.30	396.22	113594.08	56.3	25.0	
## 4	1316299	53178.55	1143.07	52035.48	56.0	25.3	
## 5	13680081	163694.73	7915.52	155779.22	239.1	87.8	
## 6	2212898	104093.67	451.78	103641.89	48.5	21.4	
##	EMP_RATE	WATER_PERC	LAND_PERC	AVG_EMP	AVG_PAY	groAvg	revAvg
## 1	0.3317675	0.033856116	0.9661438	16.19148	12.82182	1.845849	49277358
## 2	0.3635704	0.142388595	0.8576114	12.64106	19.69275	2.333000	11300000
## 3	0.3338934	0.003475910	0.9965241	16.24550	14.21126	1.874944	24865741
## 4	0.3355784	0.021494944	0.9785050	15.09711	12.22037	1.278818	8800000
## 5	0.3476897	0.048355373	0.9516447	14.97586	18.79270	3.685453	597559983
## 6	0.4047969	0.004340129	0.9956599	13.29617	19.42481	2.911014	79408163

Company Rankings

- SQL Crosstabs
- Partitioned by Industry
- Ranked by Revenue
- Ordered by Industry

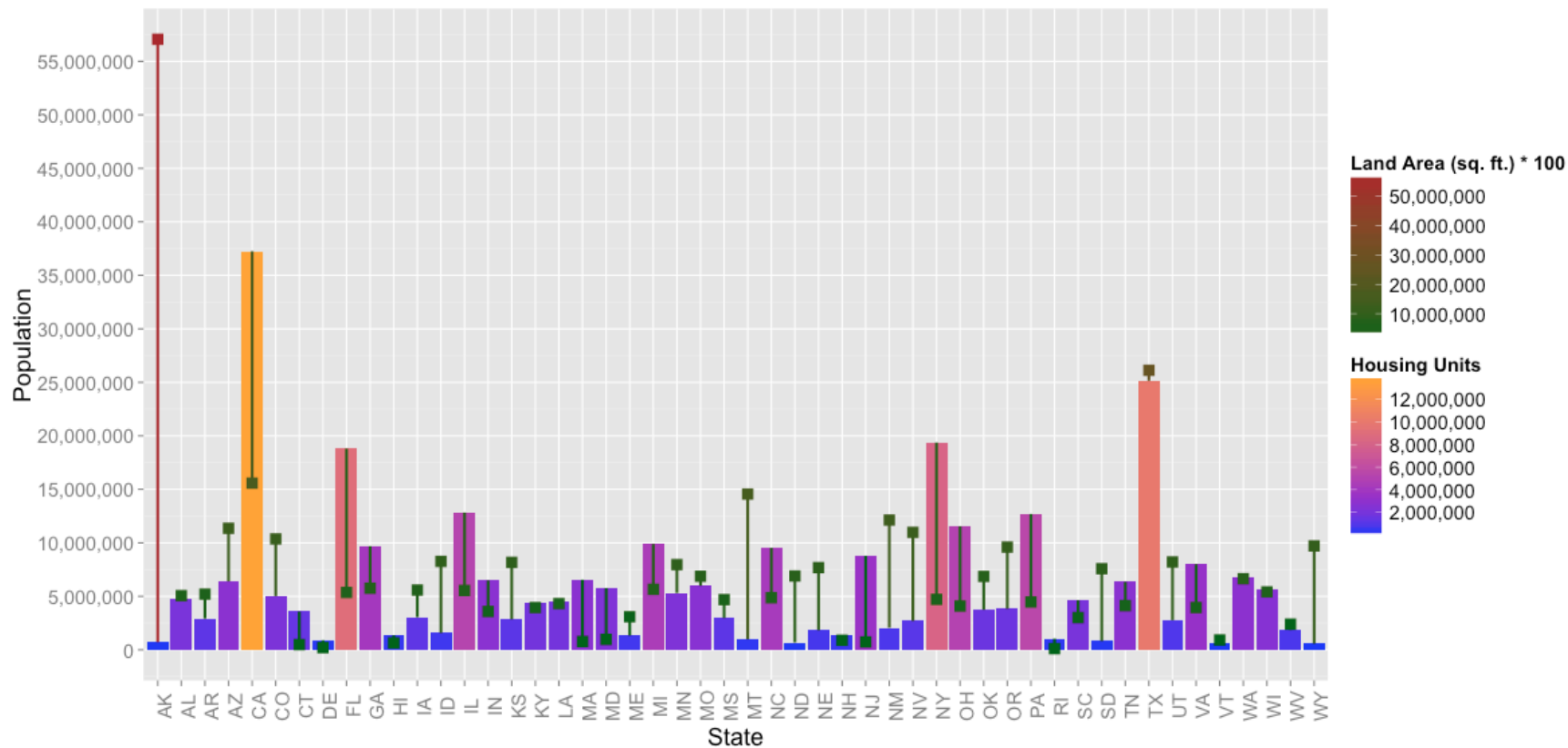
##	COMPANYID	INDUSTRY	REVENUE	INDUSTRY_RANK
## 1	4664	Advertising & Marketing	3.40549e+11	1
## 2	3033	Advertising & Marketing	3.66100e+08	2
## 3	2754	Advertising & Marketing	2.32600e+08	3
## 4	1871	Advertising & Marketing	2.08900e+08	4
## 5	1430	Advertising & Marketing	2.00900e+08	5
## 6	2010	Advertising & Marketing	1.51400e+08	6

Population Density Ranking

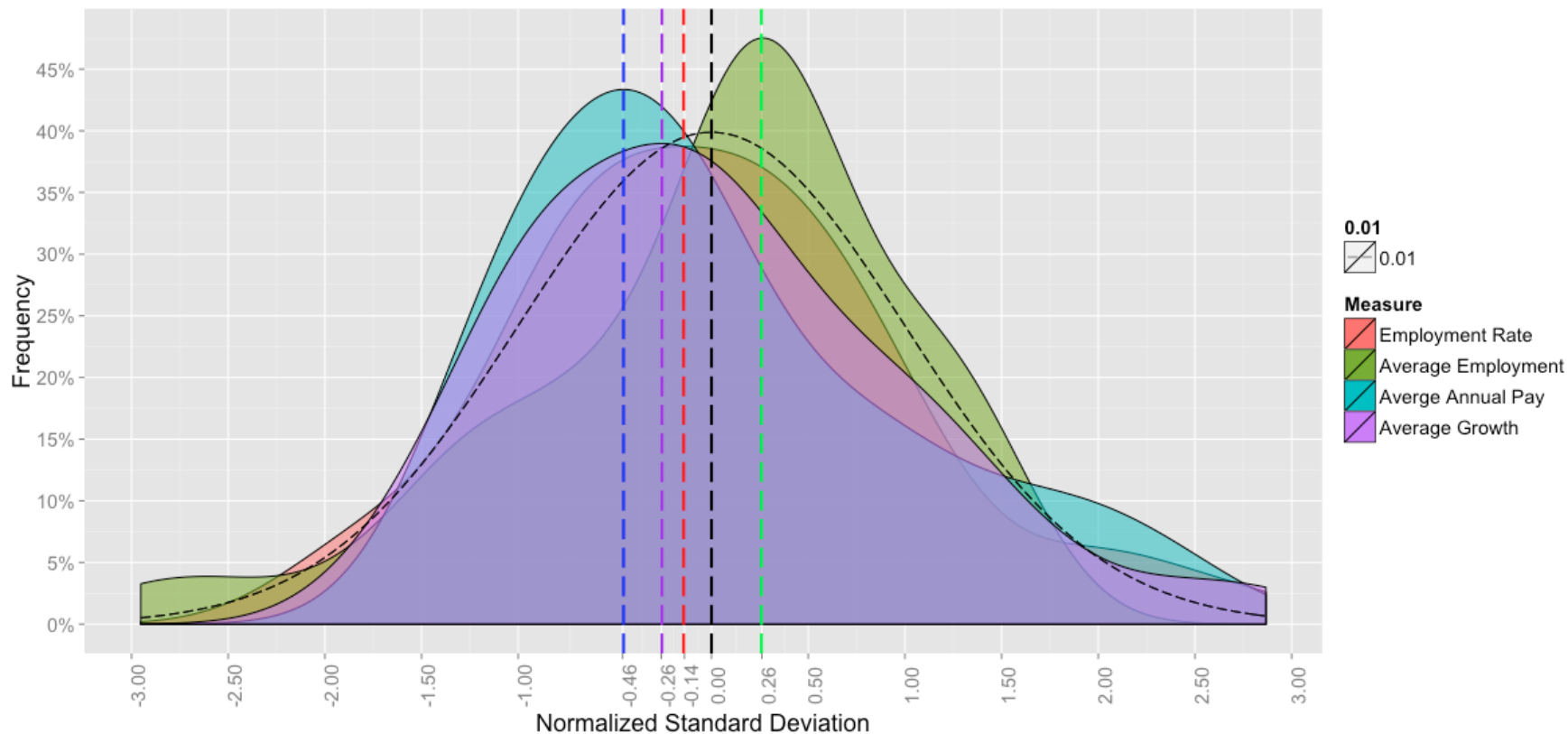
- SQL Crosstab
- Partitioned by state
- Ranked by population density

##	STATE	POPULATION	HOUSING	POPULATIONDENSITY	POPDENSITY_RANK
## 1	DC	601723	296719	9856.5	1
## 2	NJ	8791894	3553562	1195.5	2
## 3	RI	1052567	463388	1018.1	3
## 4	MA	6547629	2808254	839.4	4
## 5	CT	3574097	1487891	738.1	5
## 6	MD	5773552	2378814	594.8	6

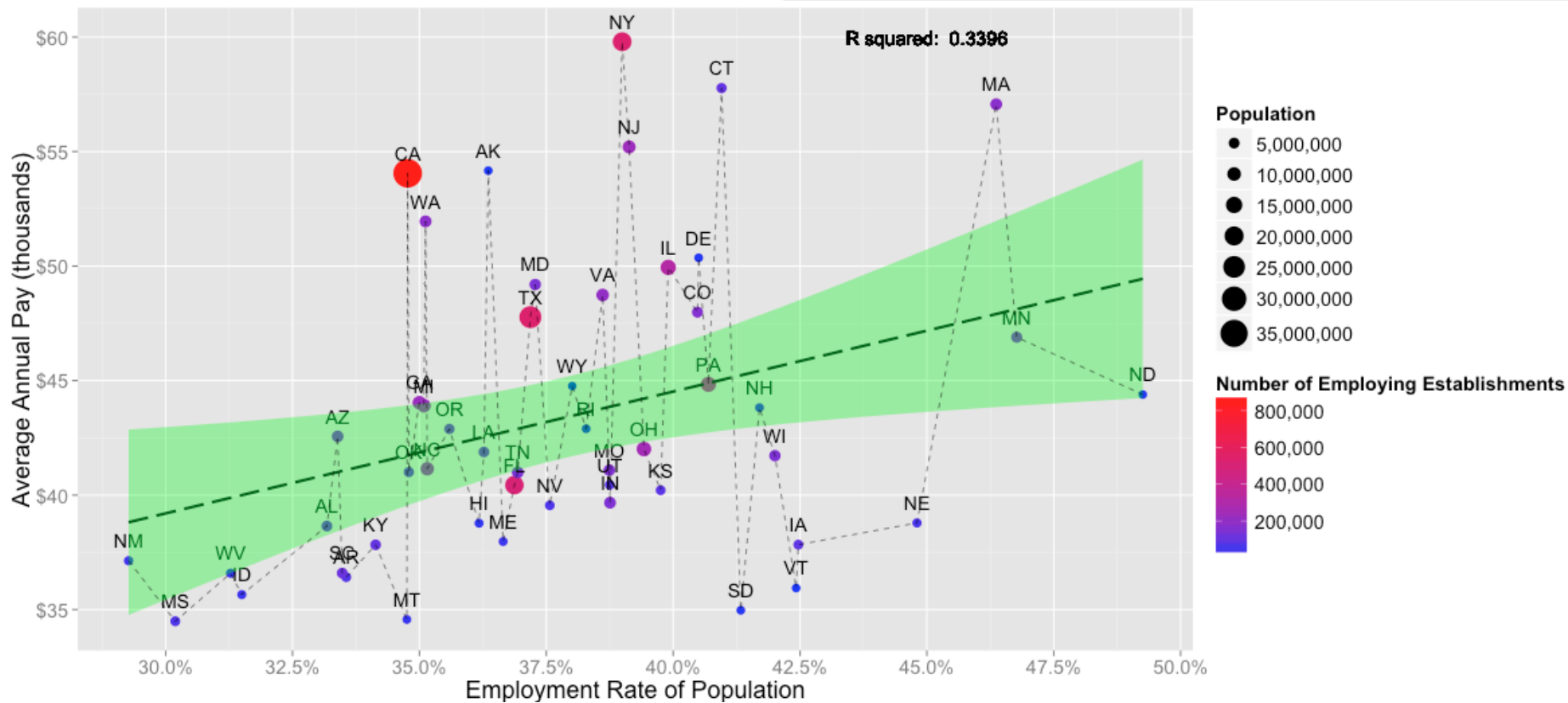
Population, Housing, and Land



Employment and Compensation

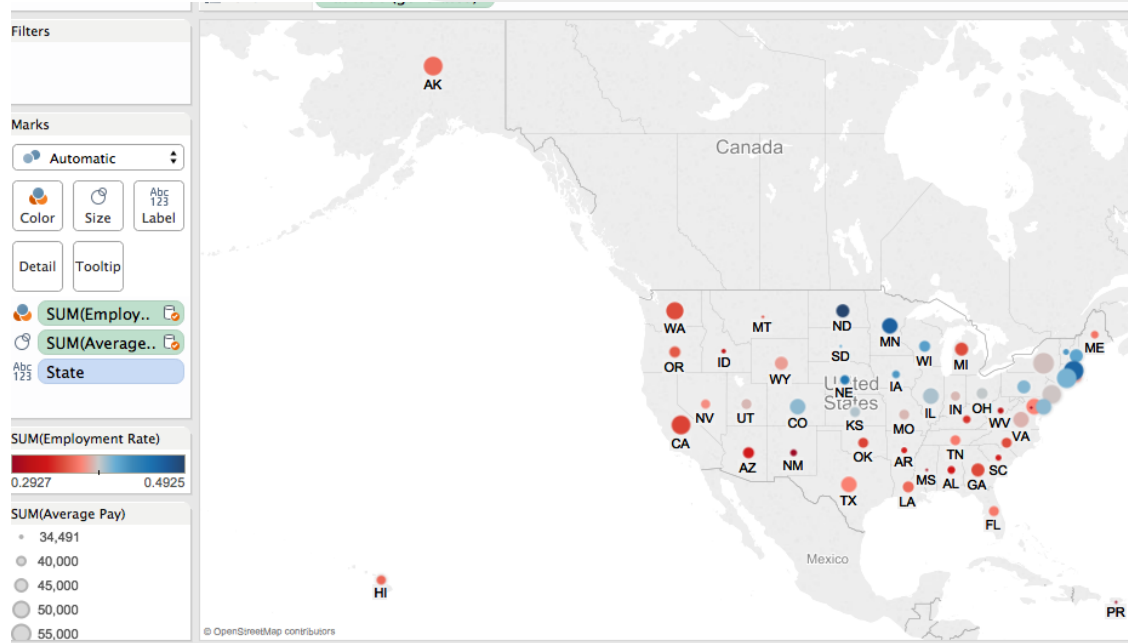


People and Money



Employment Rate and Avg Pay

- Employment Rate = $\text{Employment} / \text{Population}$
- Average Pay = $\text{Total State Payroll} / \text{Employees}$
- Size = Average Pay
- Color = Employment Rate
- Blended Data from two sources



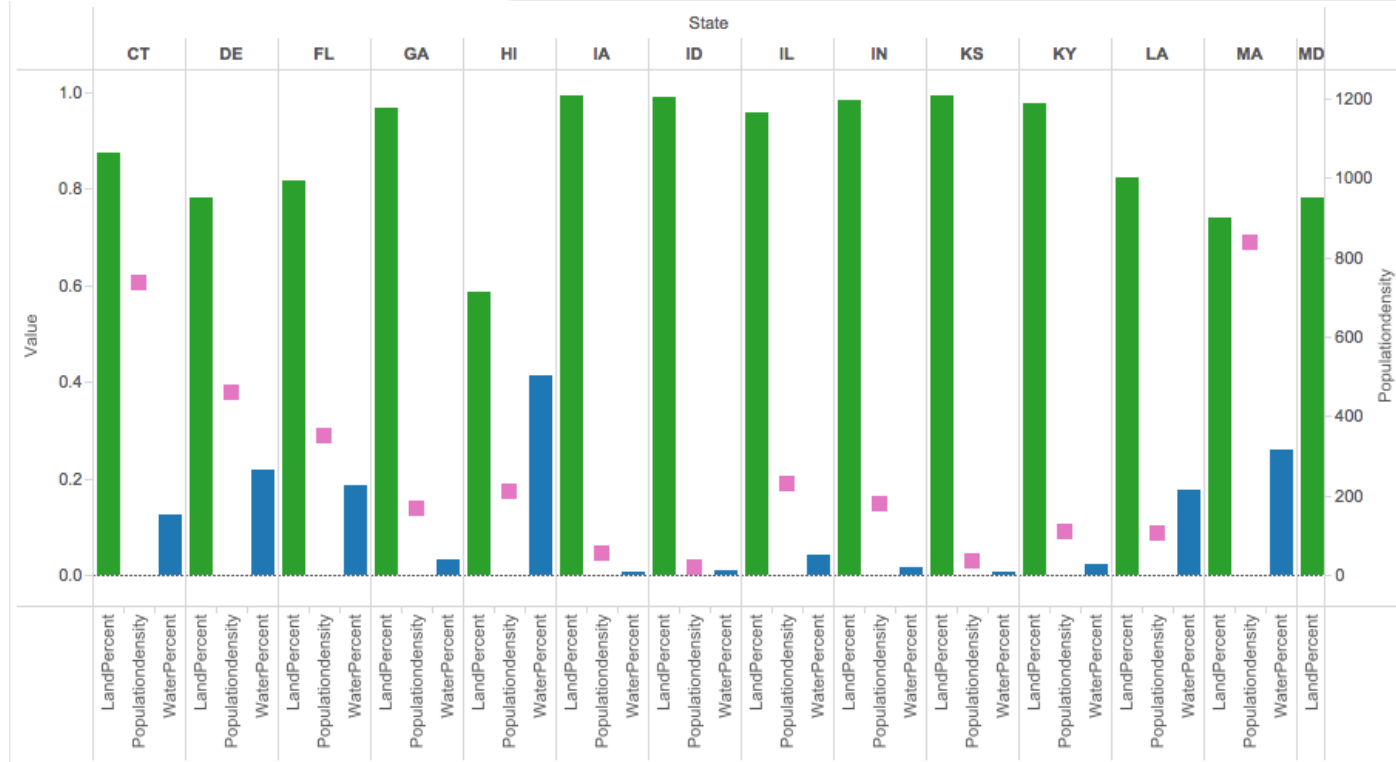
State Earnings across Industries

- Color shows the state that performs best Industry wide in terms of revenue
- Percentile shows the contribution of each industry for a singular state
- Data used from Top Companies Table

	State										
Industry	AK	AL	AR	AZ	CA	CO	CT	DC	DE	FL	GA
Advertising ..		0.12%	33.16%	8.74%	93.49%	0.91%	0.01%	24.98%	2.50%	3.82%	0.00%
Business Pr..			6.40%	9.83%	0.21%	1.49%	0.01%	3.99%	7.66%	6.62%	0.10%
Computer H..		0.87%		0.11%	1.24%	0.45%				1.03%	0.00%
Construction		39.31%	11.78%	7.58%	0.27%	2.03%	0.01%	2.72%	1.58%	4.38%	0.10%
Consumer P..		1.12%	14.98%	11.57%	0.83%	0.25%	0.00%			3.09%	0.00%
Education				3.81%	0.06%	0.42%	0.00%		3.33%	0.19%	0.00%
Energy		25.47%		0.71%	0.07%	1.14%	0.83%			3.49%	0.00%
Engineering		3.00%		1.13%	0.03%	49.75%	0.01%		6.49%	1.09%	0.00%
Environment..		1.57%		0.61%	0.20%	0.38%	98.89%		14.07%	0.34%	0.00%
Financial Ser..		0.96%		0.31%	0.29%	23.39%	0.08%			2.47%	0.00%
Food & Beve..				0.46%	0.21%	1.67%	0.00%		2.91%	1.64%	0.00%
Government ..		16.30%		1.11%	0.16%	1.05%		48.16%		2.22%	0.00%
Health	70.80%	0.96%	11.88%	3.65%	0.59%	1.89%	0.00%			7.71%	0.20%
Human Reso..		0.51%	3.62%	1.07%	0.10%	0.38%	0.00%	0.92%	33.75%	6.88%	0.00%
Insurance				0.59%	0.02%	0.80%			4.54%	7.18%	0.00%
IT Services		1.72%		21.95%	0.54%	4.21%	0.03%	6.56%	14.48%	4.49%	98.80%
Logistics & ..		0.84%	11.16%	3.01%	0.06%	0.59%	0.00%		8.70%	1.20%	0.10%
Manufacturi..		2.96%	2.58%	4.02%	0.16%	2.07%	0.01%			1.72%	0.00%
Media		0.11%		0.50%	0.07%	2.42%	0.00%	1.64%		2.91%	
Real Estate				1.01%	0.02%	0.07%		6.29%		0.03%	0.00%
Retail		1.18%	4.44%	14.70%	0.68%	0.07%	0.07%			2.20%	0.00%
Security				1.96%	0.12%	1.41%	0.01%	2.54%		0.45%	0.00%
Software	22.22%			1.94%	0.14%	0.23%	0.23%	0.23%		0.23%	

Land/Water Ratios vs Population

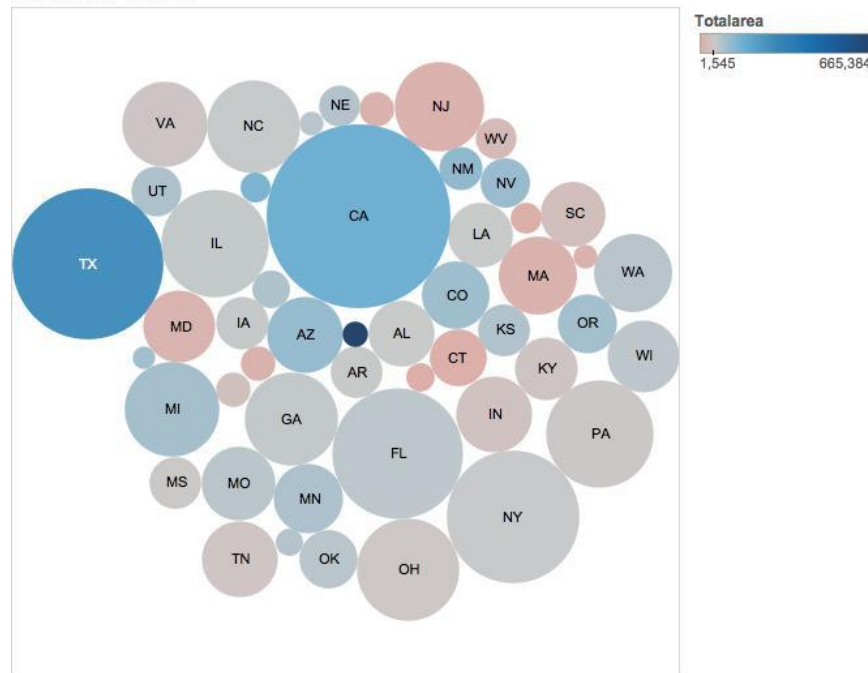
- Dual axis is used
- Double bar chart shows proportion of water vs land area per state
- Box chart shows population density



Total area with size of population

- Color reflects total area
- Size shows the population amount
- Data from a join on State Employment Data and State Housing Data

Population vs Area



State. Color shows sum of Totalarea. Size shows sum of Population. The marks are labeled by State.

Percentile Growth by City

- Parameter can be used to show nth growth per city
- Color reflects the top growth over all city
- Data from Top Companies table

City	Growth	Companyid	
Abilene	0.901	3036	0.9
Acton	9.045	238	9.0
	1.064	2727	9.0
	0.484	4063	9.0
Acworth	9.748	500	9.7
Ada	1.734	1837	1.7
Addison	4.3	639	4.3
	2.187	1463	4.3
	1.781	1784	4.3
	1.271	2392	4.3
	1.157	2574	4.3
	0.752	3352	4.3
	0.747	3368	4.3
	0.456	4122	4.3
	0.237	4637	4.3
Afton	2.648	1202	2.6
Agoura Hills	14.556	112	14.6
	0.222	4671	14.6
Akron	0.68	3534	0.7
	0.651	3622	0.7
	0.349	4410	0.7
Alachua	1.2	2502	1.2
Alameda	4.177	664	4.2

Thank you

Questions?