BGT technical document

CAVEAT I'M ONLY WORKING ON MAIN TABLE I ONLY USED A FEW COLUMNS FROM MAIN

Table of Contents

Introduction

This technical document evaluates the feasibility of using the real-time labor market information (LMI) collected by Burning Glass Technology to supplement survey and administrative data collected by federal and state governments. In contrast to designed and administrative data which have varying lag times between collection and dissemination and are often aggregated over broad occupation categories, the real-time job ads collected by Burning Glass Technology are made available within a day of the job-ad being posted and provide information at a granular level that links skill set requirements to specific occupations within broad occupation categories. In this document, the data are evaluated for use in identifying the skills sets necessary for a job in the skilled technical workforce and how these skill sets and employer demands vary within a state and across the country.

This document reviews the use of the BGT job-ads in academic research highlighting issues with the data and any validation results. It provides the results of profiling and exploratory data analyses, for both the BGT job-ads data and the Virginia Open Data/Open Jobs Data which is used to validate the BGT job-ads data. The document concludes with recommendations regarding fitness-for-use.

Burning Glass Technology

Company Background

Burning Glass Technologies is a Boston-headquartered labor market analytics firm founded in 1999, that uses artificial intelligence to collect and host a massive repository of workforce and employment data. Burning Glass collects these data primarily for commercial purposes and secondarily for research purposes. The

company markets both the data as a product as well as their consulting expertise on labor market questions to customers across a variety of industries, education institutions, local and regional governments, recruiting and staffing agencies, and corporate firms.

The company makes use of text-parsing and predictive matching methods to extract and aggregate LMI data, currently from over 40,000 sources (the number of websites has changed over the years). Data has been extracted since 2007 with the exception of 2008 and 2009. Job boards, corporate websites, and other places where job ads are posted are scanned daily and algorithms are used to identify and remove duplicate postings (close to 80% of all postings collected are marked as duplicates.) From the job-ad, the BGT software extracts the title, occupation, employer, and location, and uses natural-language technology to identify specific occupations, skills, and qualifications from the job description. Both the algorithms used to identify duplicates and scrape websites are proprietary as well as the addresses of the website that are scraped.

Review of Academic Literature

Favorite articles:

- Deming, David, and Lisa B. Kahn. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." Journal of Labor Economics 36, no. S1 (2018): S337-S369.
- Liu, Yukun. "Vacancy Postings, Skill Requirements, and the Cross-Sectional Return Predictability."
- Jeffrey Clemens, Lisa B. Kahn, and Jonathan Meer. "Dropouts Need Not Apply: The Minimum Wage and Skill Upgrading." (Sept 2018)
- Paul Mohnen, Enrico Berkes, and Bledi Taska "Can Skill Mismatch Explain Geographic and Time Variation in the Returns to College Majors? Evidence from Online Job Postings" (2017) Northwestern University
- Papageorgiou, Theodore. "Worker sorting and agglomeration economies." McGill University (2017)
- Wardrip, Keith, Stuart T. Andreason, and Mels De Zeeuw. "Uneven Opportunity: Exploring Employers' Educational. Preferences for Middle-Skills Jobs." Special Report of the Federal Reserve Banks of Philadelphia, and Atlanta (2017).
- Fuller, Joseph B., Jennifer Burrowes, Manjari Raman, Dan Restuccia, and Alexis Young. "Bridge the Gap: Rebuilding America's Middle Skills." Report, U.S. Competitiveness Project, Harvard Business School, November 2014
- Hershbein, Brad and Lisa B. Kahn (2018) Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. American Economic Review 108(7): 1737-1772. **cited all the time
- Templin, Thomas, and Lesley Hirsch. 2013. "Do Online Job Ads Predict Hiring?" New York: New York City Labor Market Information Services. ** validation example

Data Review

BGT Job-Ad Data

At the data inventory stage, potential data sources (identified in the data discovery stage) are further screened to determine if they would support the research questions. Specifically, the screening process involves assessing data sources on six factors including purpose, method, description, timeliness, selectivity, and accessibility.

- Purpose: Burning Glass collects this data primarily for commercial purposes and secondarily for research purposes. The company markets both the data as a product as well as their consulting expertise on labor market questions to customers across a variety of industries, such as higher education, local and regional government, recruiting and staffing agencies, and other corporate firms.
- Method: Burning Glass scrapes over 40,000 websites to collect data from online job postings. The data is cleaned and deduplicated to present a national view of the labor market landscape across time.
- Description: The data is centered around individual positions, with one-to-many related tables for requirements for skills, certification, and education. While the data includes various types of data, the bulk of this data is text-based.

- Timeliness: The data is collected in real-time and covers 2007 and 2010-2017. [How soon after collection is data available?]
- Selectivity: The data is intended to represent the universe of all US jobs posted online over the years mentioned above.
- Accessibility: The data is accessible via file-transfer protocol, but a data-sharing agreement may fetter its accessibility outside of the lab staff.

The structure of this data includes 8 discrete columns and 5 continuous ones.

Metric	Value
rows	63610
columns	13
discrete_columns	8
continuous_columns	5
all_missing_columns	0
total_missing_values	904
complete_rows	63158
total_observations	826930
memory_usage	9902352

Data Profiling

This section profiles the Burning Glass data, reviews its quality, and seeks to determine its useability on the following metrics.

- Completeness.
- Value validity.
- Consistency.
- Uniqueness.
- Duplication.

Here, we focus on the postings themselves which represent the central table of the data.

This first table summarizes the postings dataset. Of note, this shows that 99.2894199 % of the data are complete cases.

This second table breaks down the postings dataset by column using a superficial check for blanks. Of note, this shows that only two columns have blanks: the FIPS county and FIPS value columns are both missing 452 observations. It is important to note, however, this table qualifies missing values as blanks or NA only.

Column	# Missing	% Missing
bgtjobid	0	0.0000000
jobdate	0	0.0000000
occfam	0	0.0000000
occfamname	0	0.0000000
employer	0	0.0000000
city	0	0.0000000
state	0	0.0000000
county	0	0.0000000
fipsstate	0	0.0000000
fipscounty	452	0.0071058
fips	452	0.0071058
lat	0	0.0000000
lon	0	0.0000000

We would like also to capture invalid values in addition to the missing ones. This last table captures blanks, NA, and 'na' text values (Burning Glass entered). Here we see that more missing values are captured for the County column and new invalid values are captured for the Occupation Family code, Occupation Family Name, Employer, and City columns.

Column	Example	Class	Blanks	% Blank	NA	% NA	na
bgtjobid	38247599675	character	0	0.0	0	0.0	0
jobdate	2017-07-07	character	0	0.0	0	0.0	0
occfam	29	character	0	0.0	0	0.0	2148
occfamname	Healthcare Practitioners and Technical Occupations	character	0	0.0	0	0.0	2148
employer	na	character	0	0.0	0	0.0	13546
city	Leesburg	character	0	0.0	0	0.0	430
state	Virginia	character	0	0.0	0	0.0	0
county	Loudoun	character	452	0.7	0	0.0	0
fipsstate	51	integer	0	0.0	0	0.0	0
fipscounty	107	integer	452	0.7	452	0.7	0
fips	51107	integer	452	0.7	452	0.7	0
lat	39.0845	numeric	0	0.0	0	0.0	0
lon	-77.6066	numeric	0	0.0	0	0.0	0

We also perform a check for consistency, uniqueness, and duplication, and find that 0 have multiple dates associated with them. We can be reasonably confident the data is consistent and unique.

Exploratory Data Analysis

These histograms show the top value distribution for each column in the dataset (except the identifier column). For any variables with more than 50 possible values, only the top 20 values are shown below for clarity.

The employer is not well populated, at 20% of the values appearing at 'na.' As seen below, the 'na' values are the most common value by far, followed by Anthem Blue Cross and Booz Allen Hamilton Inc. Fairfax County is the county with the greatest number of jobs. Richmond and Arlington follow with less than half.

Throughout the month, there does appear to be a few spikes in jobs, the highest of which taking place on July 4, 2017. The next 3 days have comparable job numbers: July 25, July 21, and July 14. Throughout the month, there does appear to be a few spikes in jobs, the highest of which taking place on July 4, 2017. The next 3 days have comparable job numbers: July 25, July 21, and July 14.

It may be interesting to note that the top three coordinates given in the last two plots for the Latitude and Longitude columns, (37.5776, -77.5347), (38.8863, -77.0977), and (38.8882, -77.4552), respectively represent Richmond, Arlington (Clarendon), and Chantilly respectively.

Validation

4. BGT Job-ad Data Validation
    - Validation Results in the Literature
    - Comparison with Virginia Open Data/Open Jobs Data
    - Conclusions

FILTERING FOR VIRGINIA JULY 2017

Using the percent difference formula, as follows:

$$\frac{|V1 - V2|}{\frac{V1+V2}{2}} * 100$$

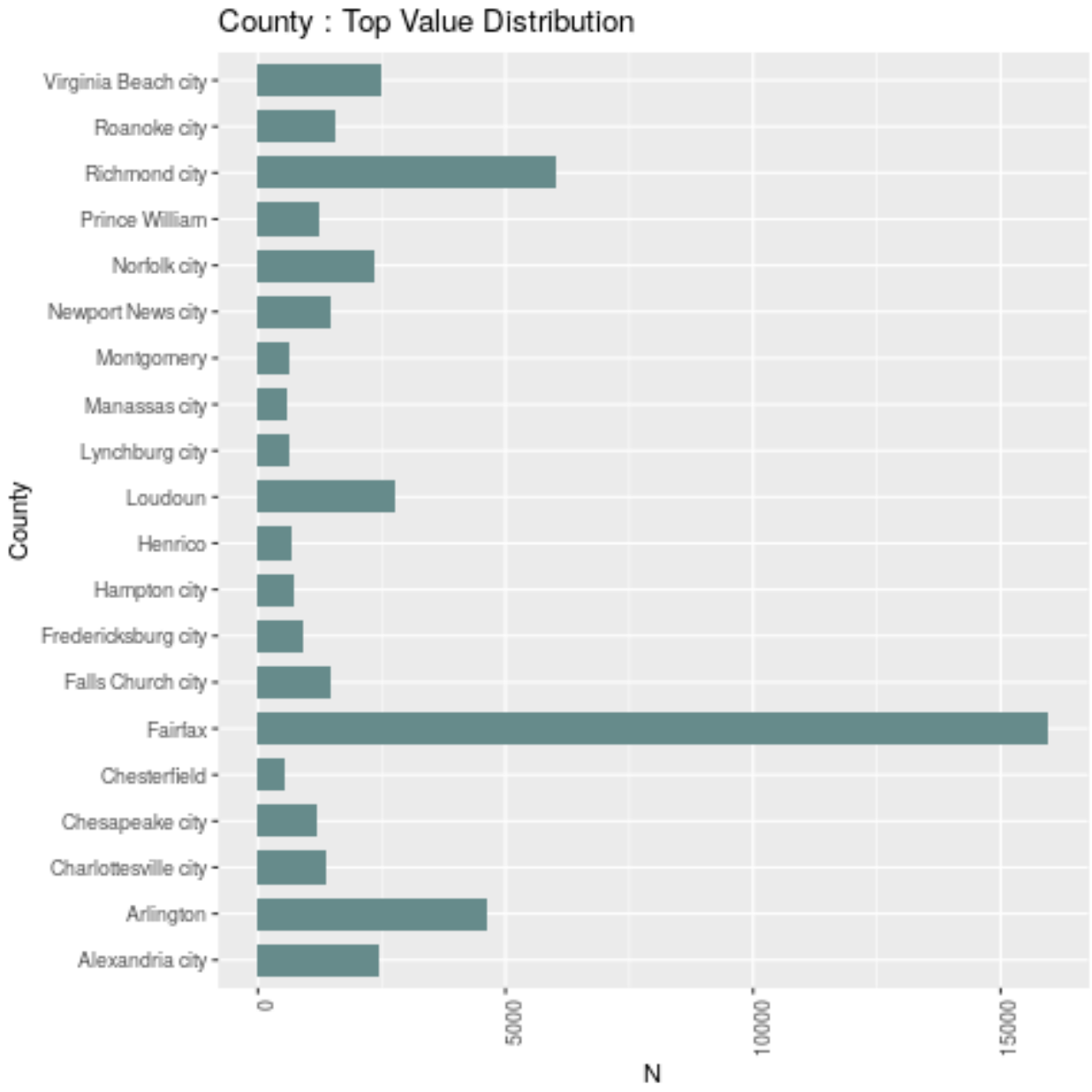

Figure 1: Figure 1. BGT Jobs in Virginia during July 2017 by County.

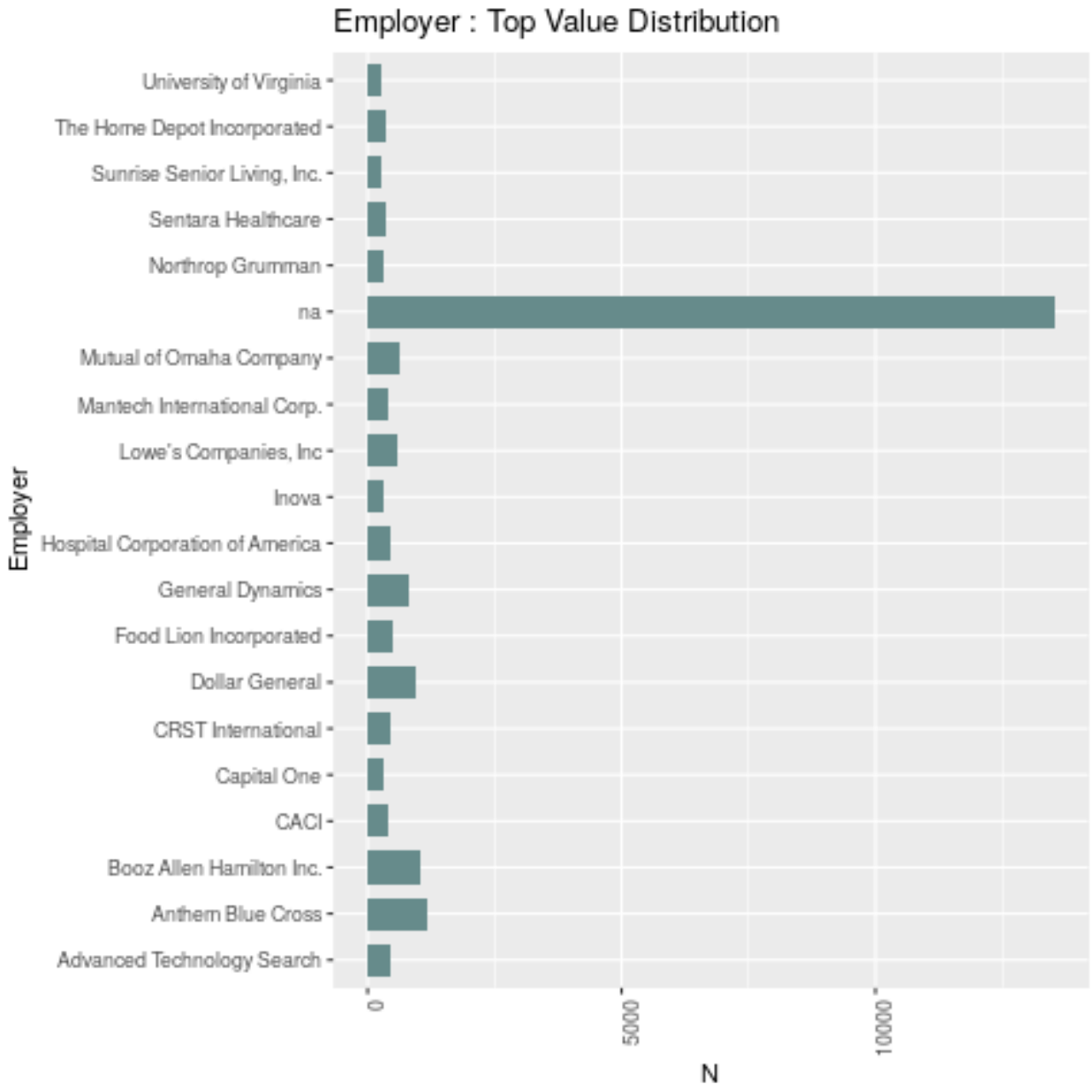

Figure 2: Figure 2. BGT Jobs in Virginia during July 2017 by Employer

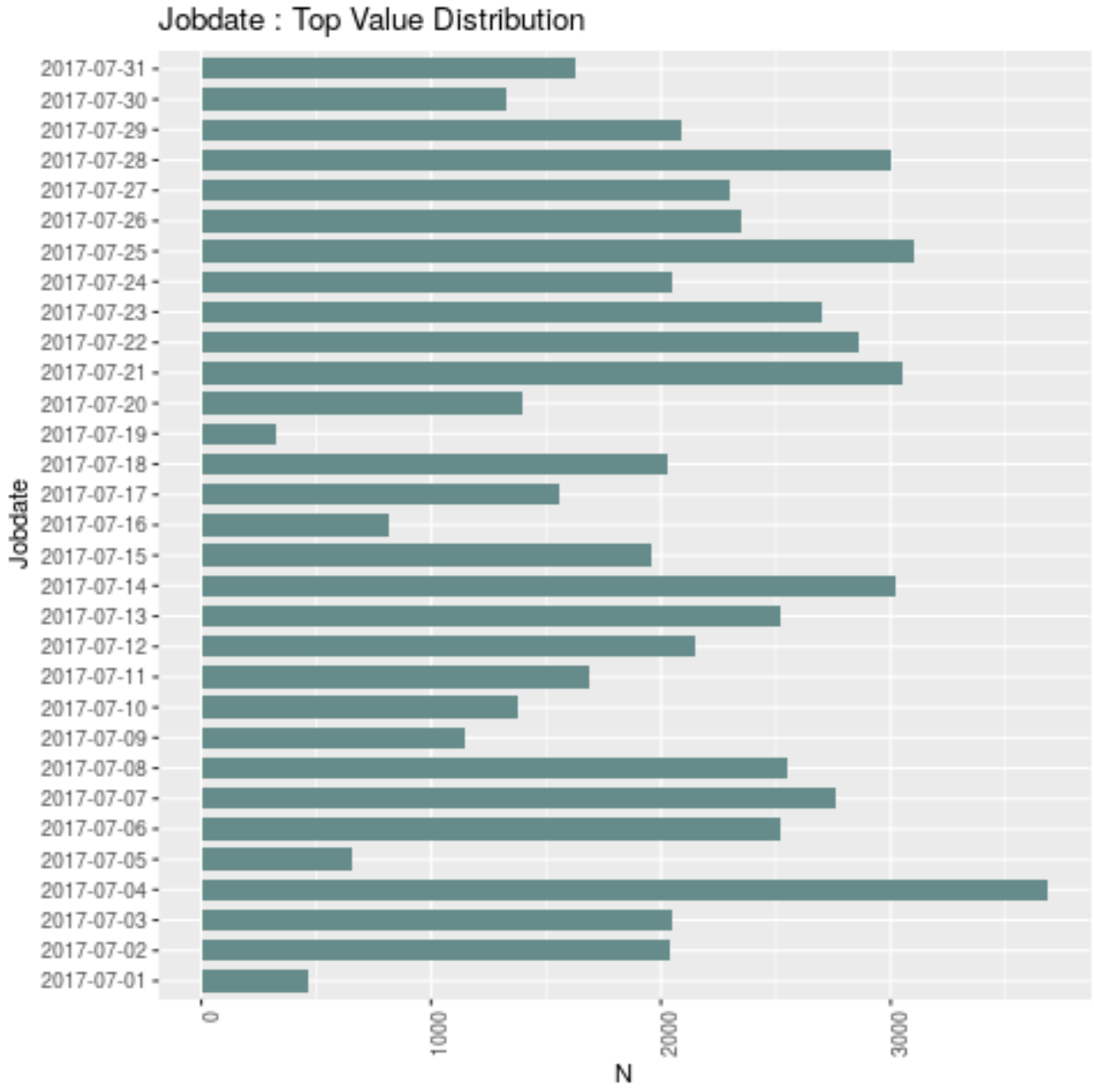

Figure 3: Figure 3. BGT Jobs in Virginia during July 2017 by Date

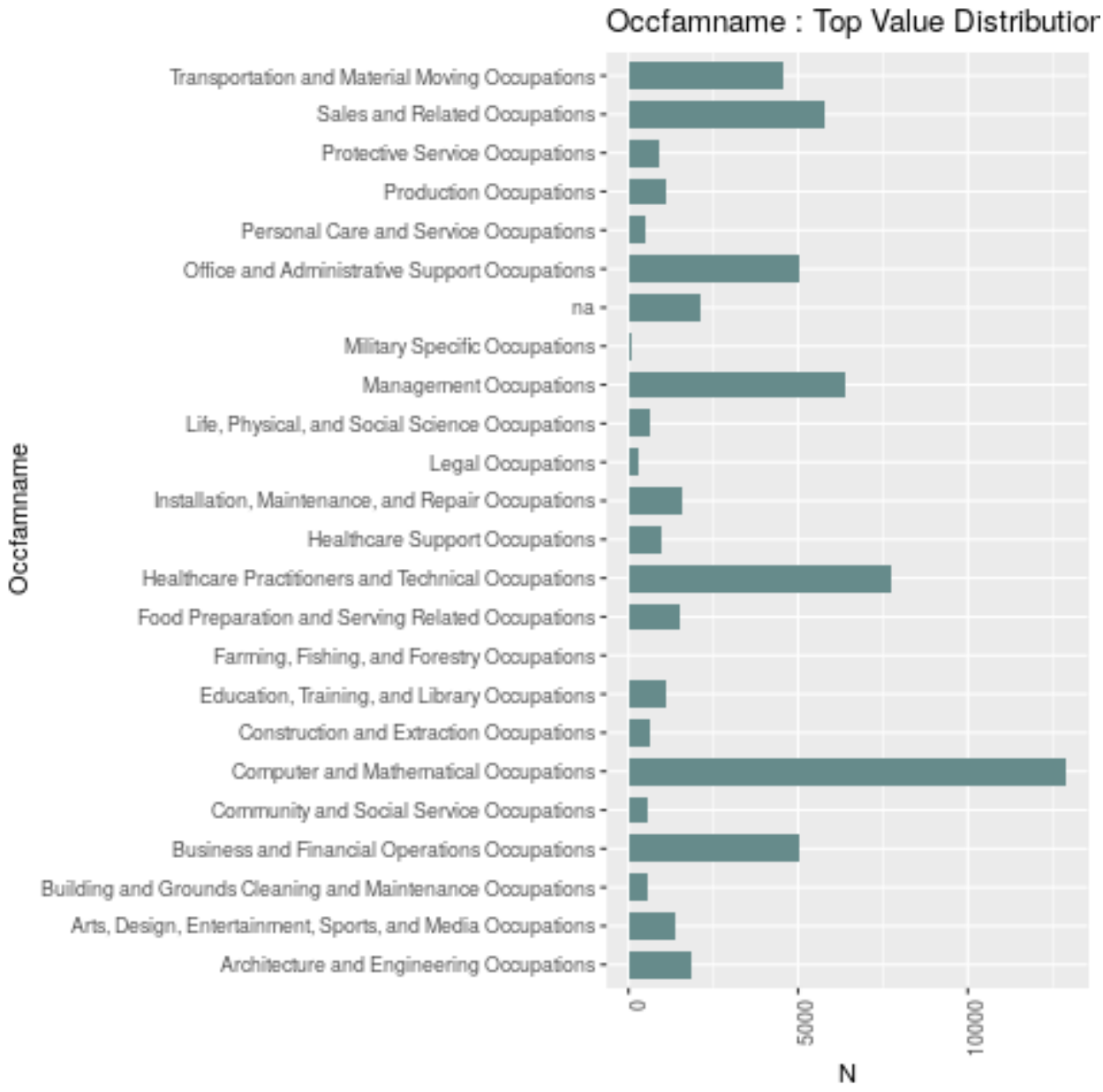

Figure 4: Figure 4. BGT Jobs in Virginia during July 2017 by Occupation Family

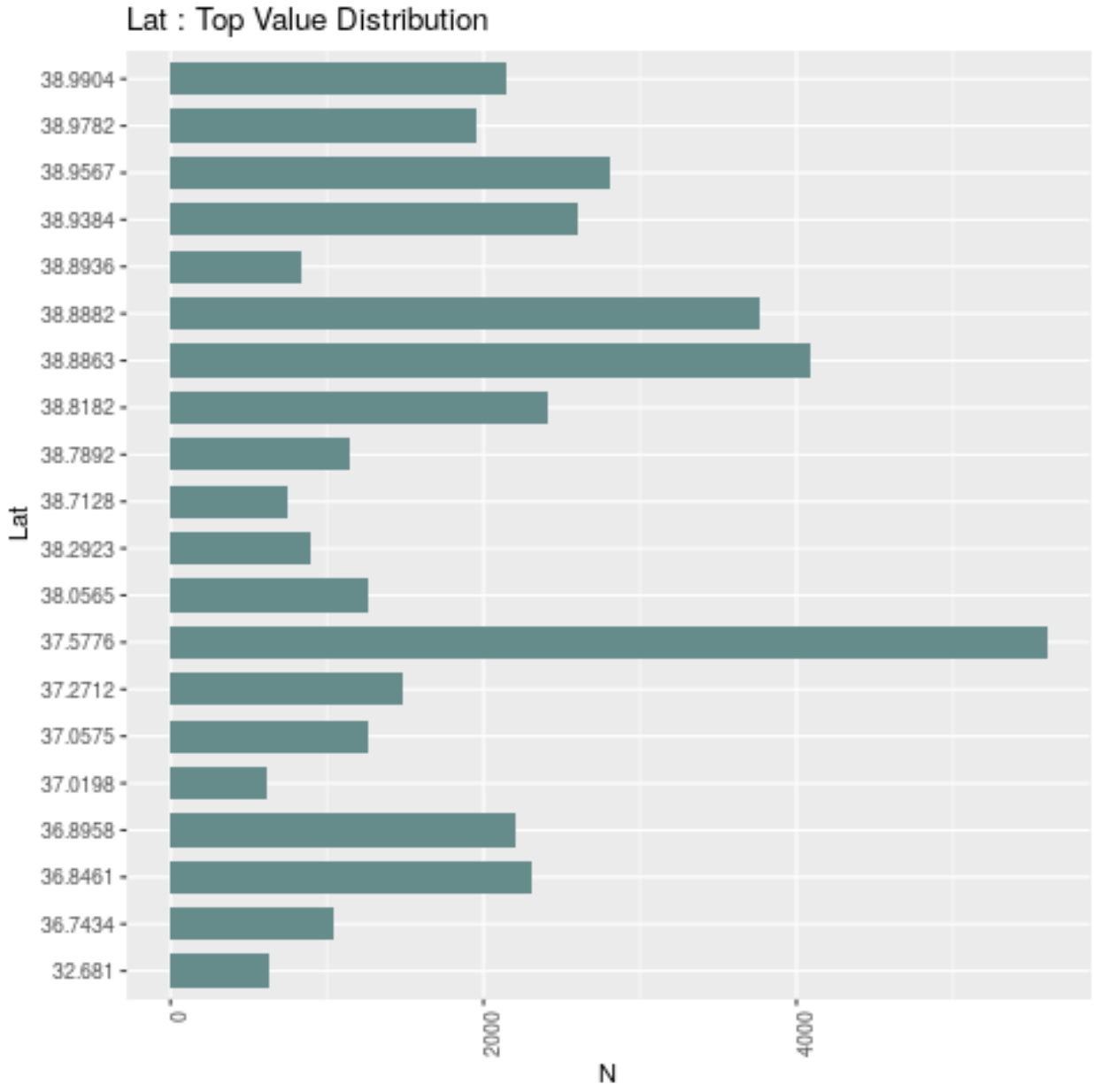

Figure 5: Figure 5. BGT Jobs in Virginia during July 2017 by Lat

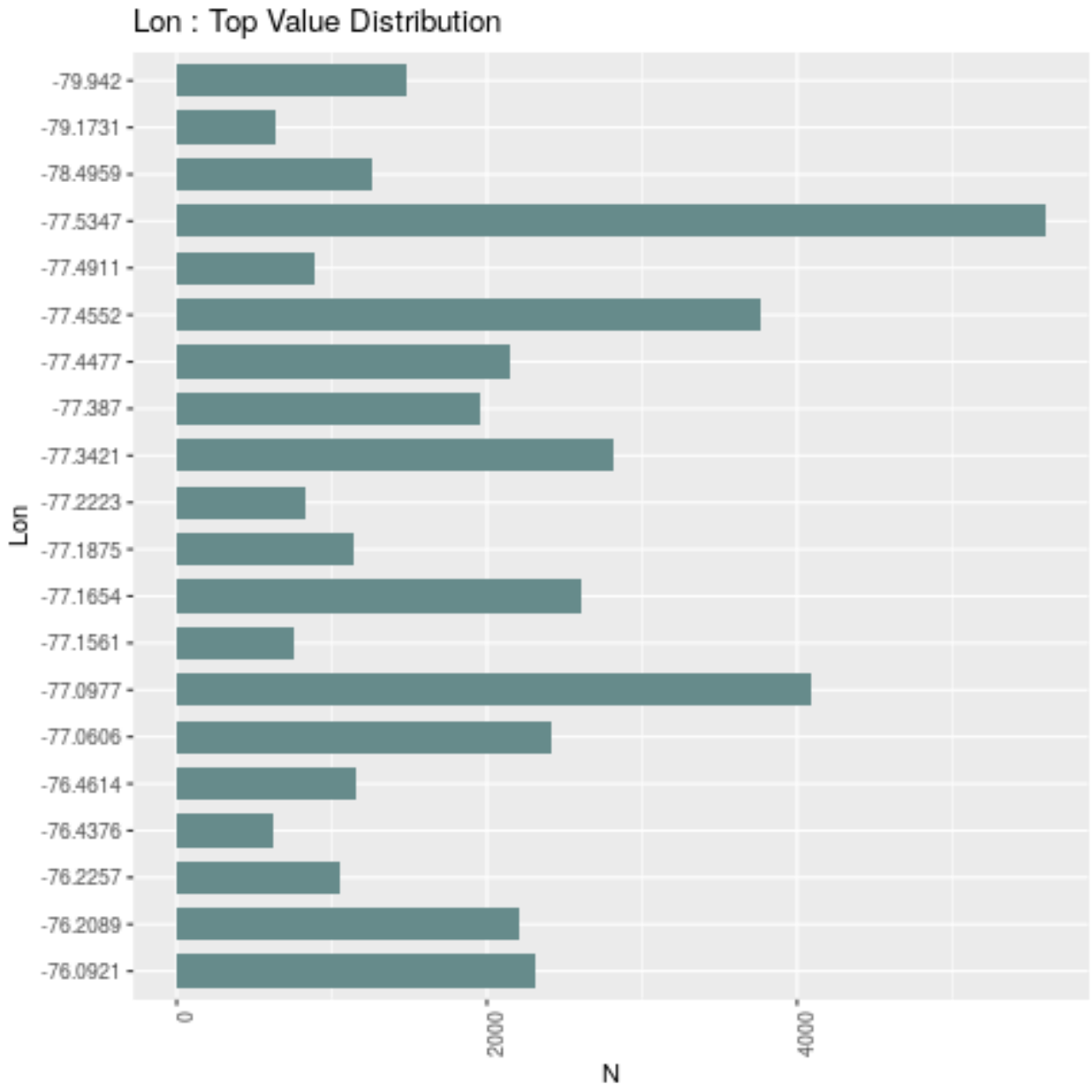

Figure 6: Figure 6. BGT Jobs in Virginia during July 2017 by :pm

The size of the two datasets differ by approximately 27.7 %.

Set	BGT Total Jobs	VA OJ Total Jobs	Difference	Percent Difference
N	63180	47807	15373	27.7

Compare by Occupation Family

occfam	occfamname	BGTcount	OJcount	Diff	PercDiff
11	Management Occupations	6408	9796	3388	41.8
13	Business and Financial Operations Occupations	5012	5701	689	12.9
15	Computer and Mathematical Occupations	12893	5313	7580	83.3
17	Architecture and Engineering Occupations	1833	1997	164	8.6
19	Life, Physical, and Social Science Occupations	613	220	393	94.4
21	Community and Social Service Occupations	597	191	406	103.0
23	Legal Occupations	290	88	202	106.9
25	Education, Training, and Library Occupations	1110	571	539	64.1
27	Arts, Design, Entertainment, Sports, and Media Occupations	1369	880	489	43.5
29	Healthcare Practitioners and Technical Occupations	7743	3277	4466	81.1
31	Healthcare Support Occupations	972	416	556	80.1
33	Protective Service Occupations	872	508	364	52.8
35	Food Preparation and Serving Related Occupations	1518	593	925	87.6
37	Building and Grounds Cleaning and Maintenance Occupations	577	242	335	81.8
39	Personal Care and Service Occupations	490	130	360	116.1
41	Sales and Related Occupations	5771	1716	4055	108.3
43	Office and Administrative Support Occupations	5055	2858	2197	55.5
45	Farming, Fishing, and Forestry Occupations	34	NA	NA	NA
47	Construction and Extraction Occupations	599	306	293	64.8
49	Installation, Maintenance, and Repair Occupations	1563	1112	451	33.7
51	Production Occupations	1079	397	682	92.4
53	Transportation and Material Moving Occupations	4542	3235	1307	33.6
55	Military Specific Occupations	107	NA	NA	NA

Compare by Economic Region

Region	GOorg	BGTcount	OJcount	Diff	PercDiff
1	UVA Wise	1204	1132	72	6.2
2	VT Economic Development	3688	4141	453	11.6
3	Southside Planning District Commission	1547	1509	38	2.5
4	GROW Capital Jobs Foundation	9071	6914	2157	27.0
5	Reinvent Hampton Roads	10126	8350	1776	19.2
6	George Washington Regional Commission	2168	1193	975	58.0
7	Northern Virginia Regional Council	28690	20745	7945	32.1
8	Northern Shenandoah Valley Regional Commission	2565	1964	601	26.5
9	Central Virginia Partnership for Economic Developmet	2807	1859	948	40.6
NA	NA	1314	NA	NA	NA

Compare by Geography

The distribution of jobs throughout Virginia appears as expected, with coverage throughout all regions of

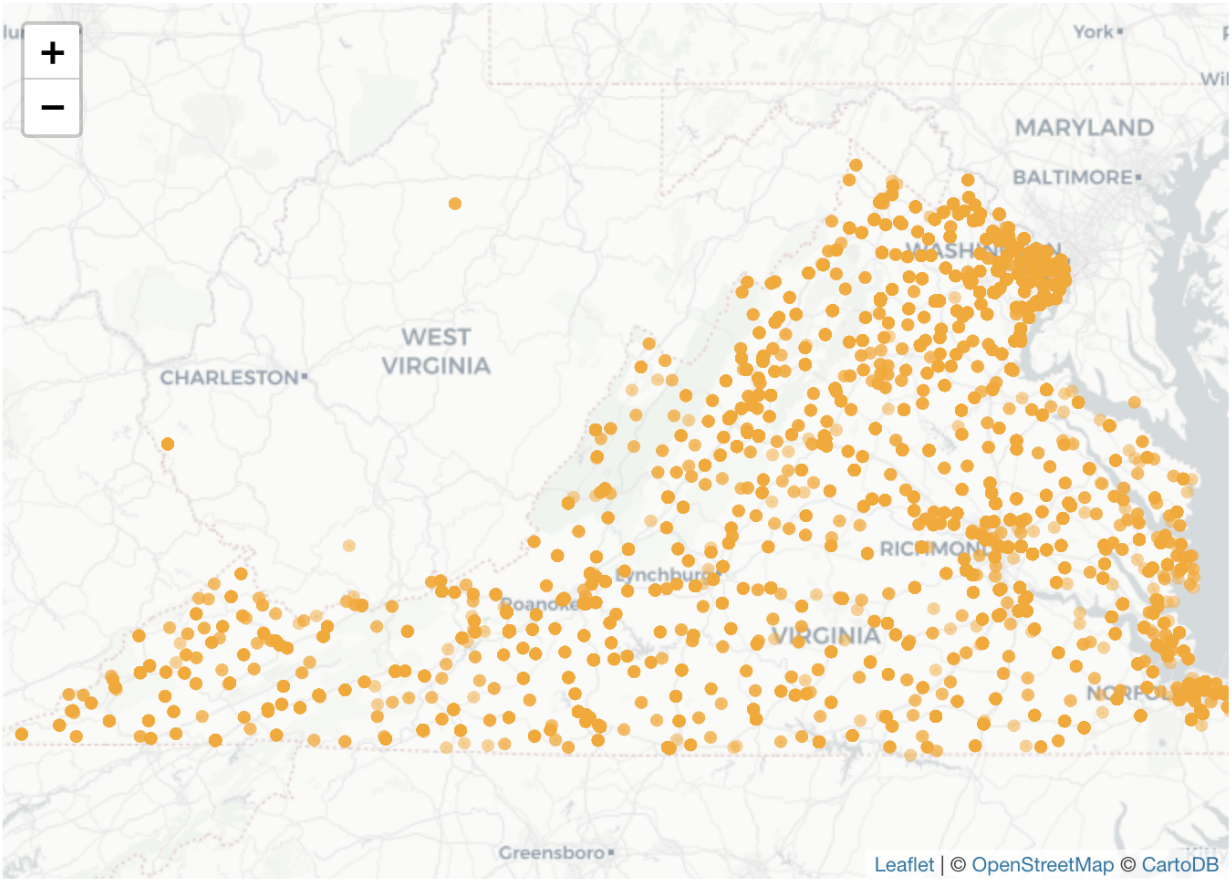

Figure 7: Figure 7. BGT Job Locations in Virginia during July 2017 by Lat

Virginia and density surrounding the major urban areas. Notably, the increased density aligns with the areas around the DC metro area, the Norfolk/Virginia Beach area, and the city of Richmond.

If we look at this same map in aggregate across counties, we see the same density patterns (darker reds indicating greater density) except Blacksburg and Charlottesville also become notable areas of density.

APPENDIX

Compare by Economic Region & County

Region	GOorg	County	BGTcount	OJcount
1	UVA Wise	Bland County	15	11
1	UVA Wise	Bristol city	119	434
1	UVA Wise	Buchanan County	48	15
1	UVA Wise	Carroll County	49	15
1	UVA Wise	Dickenson County	30	13
1	UVA Wise	Galax city	76	67
1	UVA Wise	Grayson County	27	15
1	UVA Wise	Lee County	60	22
1	UVA Wise	Norton city	34	34
1	UVA Wise	Russell County	51	45

Region	GOorg	County	BGTcount	OJcount
1	UVA Wise	Scott County	57	24
1	UVA Wise	Smyth County	103	39
1	UVA Wise	Tazewell County	196	149
1	UVA Wise	Washington County	51	87
1	UVA Wise	Wise County	108	103
1	UVA Wise	Wythe County	180	59
2	VT Economic Development	Alleghany County	54	58
2	VT Economic Development	Amherst County	63	31
2	VT Economic Development	Appomattox County	27	10
2	VT Economic Development	Bedford County	176	103
2	VT Economic Development	Botetourt County	114	31
2	VT Economic Development	Campbell County	95	73
2	VT Economic Development	Craig County	16	NA
2	VT Economic Development	Floyd County	30	6
2	VT Economic Development	Franklin County	179	92
2	VT Economic Development	Giles County	51	29
2	VT Economic Development	Lynchburg city	22	780
2	VT Economic Development	Montgomery County	656	788
2	VT Economic Development	Pulaski County	219	676
2	VT Economic Development	Radford city	1	NA
2	VT Economic Development	Roanoke city	1574	1264
2	VT Economic Development	Roanoke County	47	18
2	VT Economic Development	Salem city	364	182
3	Southside Planning District Commission	Amelia County	20	9
3	Southside Planning District Commission	Brunswick County	58	17
3	Southside Planning District Commission	Buckingham County	27	19
3	Southside Planning District Commission	Charlotte County	63	NA
3	Southside Planning District Commission	Cumberland County	11	7
3	Southside Planning District Commission	Danville city	463	527
3	Southside Planning District Commission	Halifax County	119	407
3	Southside Planning District Commission	Henry County	77	43
3	Southside Planning District Commission	Lunenburg County	27	16
3	Southside Planning District Commission	Martinsville city	219	149
3	Southside Planning District Commission	Mecklenburg County	162	121
3	Southside Planning District Commission	Nottoway County	62	116
3	Southside Planning District Commission	Patrick County	37	7
3	Southside Planning District Commission	Pittsylvania County	72	40
3	Southside Planning District Commission	Prince Edward County	130	31
4	GROW Capital Jobs Foundation	Charles City County	11	NA
4	GROW Capital Jobs Foundation	Chesterfield County	560	440
4	GROW Capital Jobs Foundation	Dinwiddie County	463	74
4	GROW Capital Jobs Foundation	Emporia city	58	114
4	GROW Capital Jobs Foundation	Goochland County	207	90
4	GROW Capital Jobs Foundation	Greensville County	25	5
4	GROW Capital Jobs Foundation	Hanover County	438	271
4	GROW Capital Jobs Foundation	Henrico County	707	463
4	GROW Capital Jobs Foundation	Hopewell city	110	56
4	GROW Capital Jobs Foundation	New Kent County	95	12
4	GROW Capital Jobs Foundation	Petersburg city	8	276
4	GROW Capital Jobs Foundation	Powhatan County	44	90
4	GROW Capital Jobs Foundation	Prince George County	158	132
4	GROW Capital Jobs Foundation	Richmond city	6129	4759

Region	GOorg	County	BGTcount	OJcount
4	GROW Capital Jobs Foundation	Surry County	29	17
4	GROW Capital Jobs Foundation	Sussex County	29	37
5	Reinvent Hampton Roads	Accomack County	246	206
5	Reinvent Hampton Roads	Chesapeake city	1201	1131
5	Reinvent Hampton Roads	Franklin city	86	176
5	Reinvent Hampton Roads	Hampton city	87	610
5	Reinvent Hampton Roads	Isle of Wight County	90	402
5	Reinvent Hampton Roads	James City County	427	11
5	Reinvent Hampton Roads	Newport News city	2104	1777
5	Reinvent Hampton Roads	Norfolk city	2358	1230
5	Reinvent Hampton Roads	Northampton County	58	14
5	Reinvent Hampton Roads	Poquoson city	24	5
5	Reinvent Hampton Roads	Portsmouth city	447	421
5	Reinvent Hampton Roads	Southampton County	46	12
5	Reinvent Hampton Roads	Suffolk city	328	235
5	Reinvent Hampton Roads	Virginia Beach city	2482	1656
5	Reinvent Hampton Roads	Williamsburg city	58	387
5	Reinvent Hampton Roads	York County	84	77
6	George Washington Regional Commission	Caroline County	75	36
6	George Washington Regional Commission	Essex County	77	19
6	George Washington Regional Commission	Fredericksburg city	891	511
6	George Washington Regional Commission	Gloucester County	105	85
6	George Washington Regional Commission	King and Queen County	277	5
6	George Washington Regional Commission	King George County	73	156
6	George Washington Regional Commission	King William County	47	26
6	George Washington Regional Commission	Lancaster County	42	27
6	George Washington Regional Commission	Mathews County	79	8
6	George Washington Regional Commission	Middlesex County	40	21
6	George Washington Regional Commission	Northumberland County	28	16
6	George Washington Regional Commission	Richmond County	23	9
6	George Washington Regional Commission	Spotsylvania County	49	31
6	George Washington Regional Commission	Stafford County	307	227
6	George Washington Regional Commission	Westmoreland County	55	16
7	Northern Virginia Regional Council	Alexandria city	2422	NA
7	Northern Virginia Regional Council	Arlington County	4637	3053
7	Northern Virginia Regional Council	Fairfax County	15986	11162
7	Northern Virginia Regional Council	Falls Church city	28	1084
7	Northern Virginia Regional Council	Loudoun County	3782	2655
7	Northern Virginia Regional Council	Manassas city	582	436
7	Northern Virginia Regional Council	Manassas Park city	14	NA
7	Northern Virginia Regional Council	Prince William County	1239	1216
8	Northern Shenandoah Valley Regional Commission	Augusta County	250	98
8	Northern Shenandoah Valley Regional Commission	Bath County	55	18
8	Northern Shenandoah Valley Regional Commission	Buena Vista city	21	7
8	Northern Shenandoah Valley Regional Commission	Clarke County	86	48
8	Northern Shenandoah Valley Regional Commission	Frederick County	102	25
8	Northern Shenandoah Valley Regional Commission	Harrisonburg city	446	260
8	Northern Shenandoah Valley Regional Commission	Highland County	16	26
8	Northern Shenandoah Valley Regional Commission	Lexington city	86	74
8	Northern Shenandoah Valley Regional Commission	Page County	105	54
8	Northern Shenandoah Valley Regional Commission	Rockbridge County	40	14
8	Northern Shenandoah Valley Regional Commission	Rockingham County	261	110

Region	GOorg	County	BGTcount	OJcount
8	Northern Shenandoah Valley Regional Commission	Shenandoah County	171	163
8	Northern Shenandoah Valley Regional Commission	Staunton city	127	184
8	Northern Shenandoah Valley Regional Commission	Warren County	135	98
8	Northern Shenandoah Valley Regional Commission	Waynesboro city	138	96
8	Northern Shenandoah Valley Regional Commission	Winchester city	526	689
9	Central Virginia Partnership for Economic Developmet	Albemarle County	87	13
9	Central Virginia Partnership for Economic Developmet	Charlottesville city	1320	946
9	Central Virginia Partnership for Economic Developmet	Culpeper County	214	175
9	Central Virginia Partnership for Economic Developmet	Fauquier County	447	235
9	Central Virginia Partnership for Economic Developmet	Fluvanna County	57	367
9	Central Virginia Partnership for Economic Developmet	Greene County	48	26
9	Central Virginia Partnership for Economic Developmet	Louisa County	99	32
9	Central Virginia Partnership for Economic Developmet	Madison County	64	11
9	Central Virginia Partnership for Economic Developmet	Nelson County	54	6
9	Central Virginia Partnership for Economic Developmet	Orange County	71	27
9	Central Virginia Partnership for Economic Developmet	Rappahannock County	346	21
NA	NA	NA	1314	NA

Post Script? Questions Remaining for Burning Glass

- If occfam is missing, is occfamname always missing as well?
- what naming standardizations have been done to the employer column?
- How does burning glass handle subsidiaries?

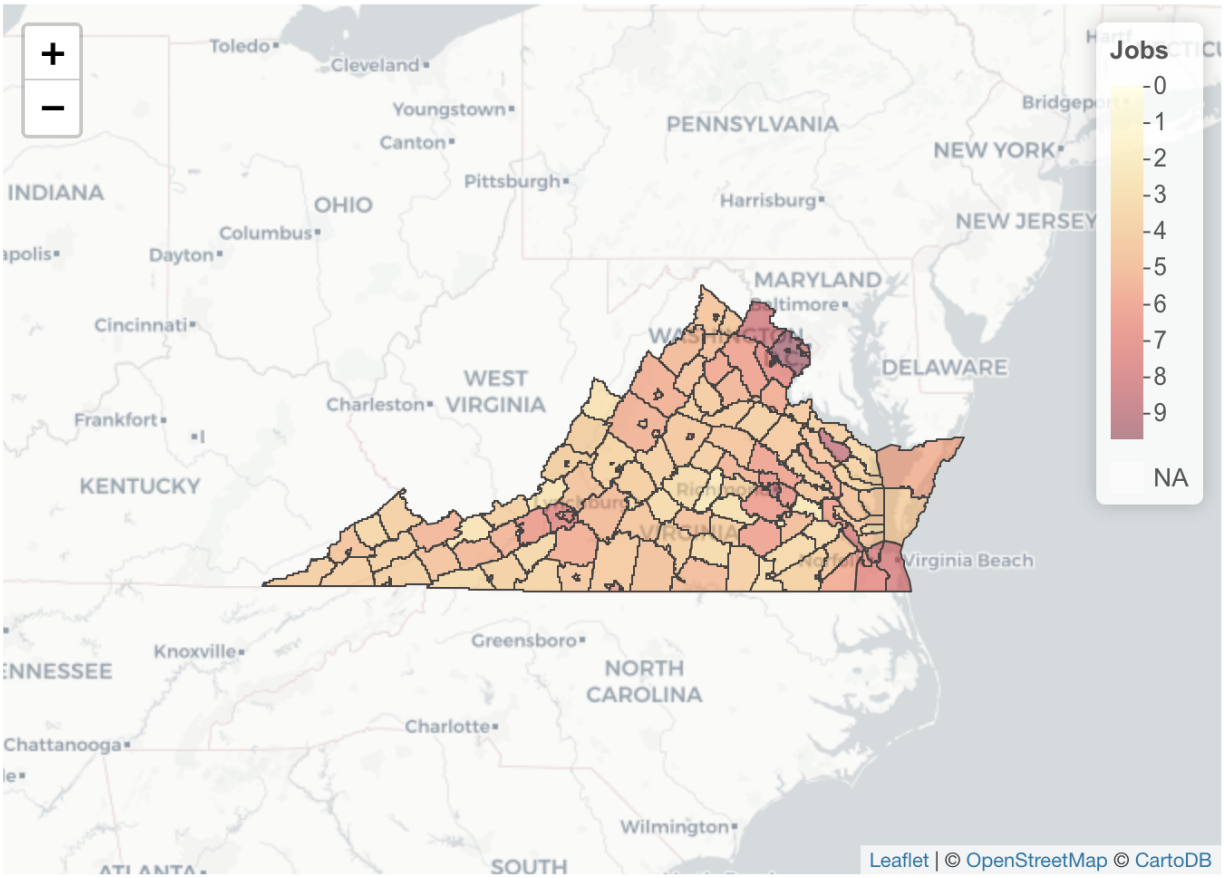

Figure 8: Figure 8. BGT Job Density in Virginia Counties during July 2017 by :pm