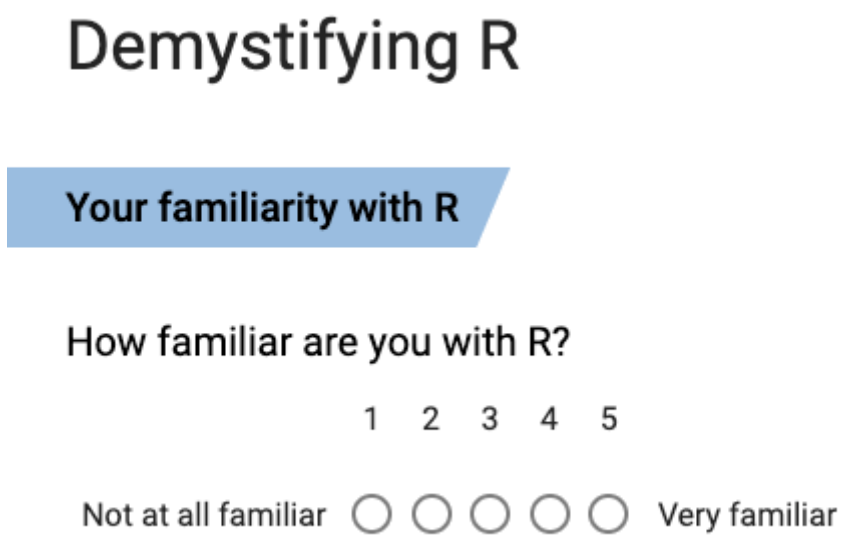
# Demystifying R: A Guided Tour

David Keyes
R for the Rest of Us

# Before We Start

Please take the survey at https://rfortherestofus.com/survey/



# Logistics

Everything will be posted at https://rfor.us/demystifying-feb-2021

If you have any questions, please put them in the chat (I'll stop if necessary)

There will also be time for Q&A at the end

# Who Am I?

# Getting Started with R

# What is R?

# Download and Install R

The first thing you need to do is download the R software. Go to the Comprehensive R Archive Network (aka “CRAN”) website and download the software for your operating system (Windows, Mac, or Linux).

# R

# RStudio

| R: Engine | RStudio: Dashboard |
| --- | --- |
|  |  |

Courtesy Modern Dive

# Download and Install RStudio

Download RStudio at the RStudio website. Ignore the various versions listed there. All you need is the latest version of RStudio Desktop.

# RStudio

# Packages

# Packages

| R: A new phone | R Packages: Apps you can download |
| --- | --- |
|  |   |

Courtesy Modern Dive

# Examples of Packages

Tidyverse

Packages Articles Learn Help Contribute


## R packages for data science

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

# Examples of Packages

## gendercodeR

The goal of gendercodeR is to allow simple recoding of freetext gender responses.

### Why would we do this?

Researchers who collect self-reported demographic data from respondents occasionally collect gender using a free-text response option. This has the advantage of respecting the gender diversity of respondents without prompting users and potentially including misleading responses. However, this presents a challenge to researchers in that some inconsistencies in typography and spelling create a larger set of responses than would be required to fully capture the demographic characteristics of the sample.

For example, male participants may provide freetext responses as "male", "man", "mail", "mael". Non-binary participants may provide responses as "nonbinary", "enby", "non-binary", "non binary"

This package uses dictionaries of common misspellings to recode these freetext responses into a consistent set of responses.

# Why Use R?

# Data Analysis in a Snap

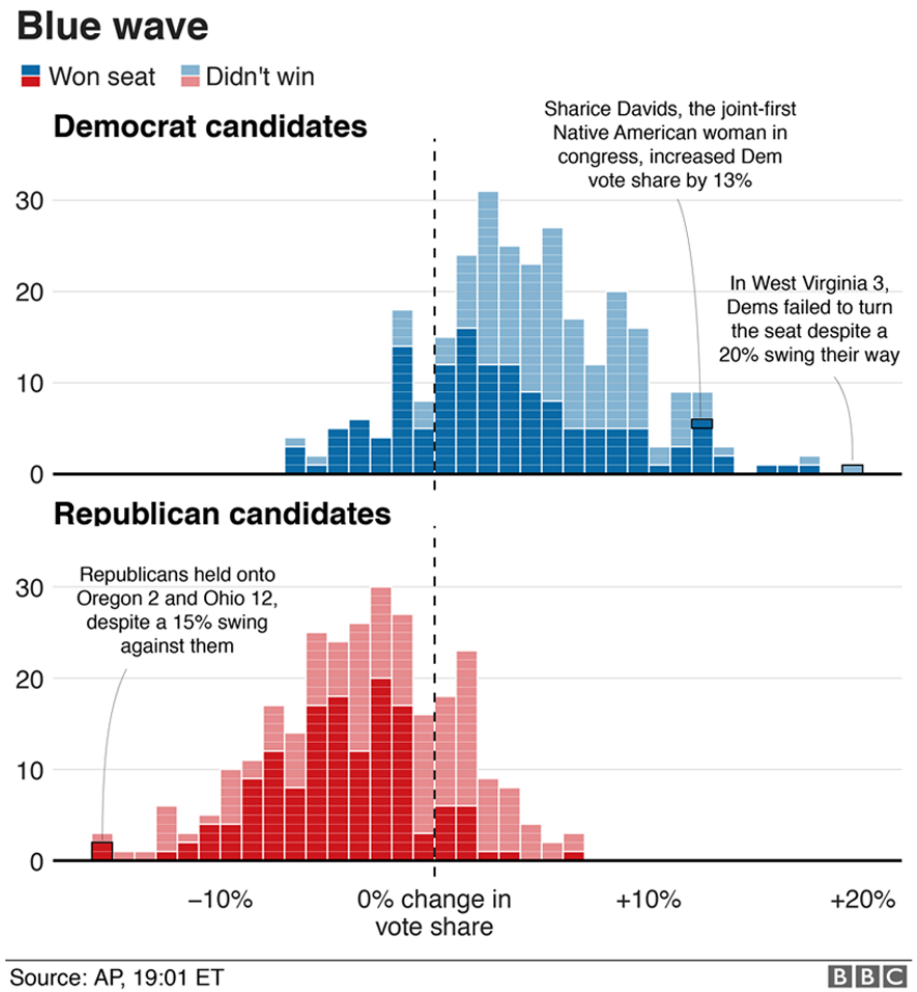
| gender | education | marital_status | height |
|---|---|---|---|
| male | High School | Married | 164.7 |
| male | High School | Married | 164.7 |
| male | High School | Married | 164.7 |
| male | NA | NA | 105.4 |
| female | Some College | LivePartner | 168.4 |
| male | NA | NA | 133.1 |
| male | NA | NA | 130.6 |
| female | College Grad | Married | 166.7 |
| female | College Grad | Married | 166.7 |
| female | College Grad | Married | 166.7 |

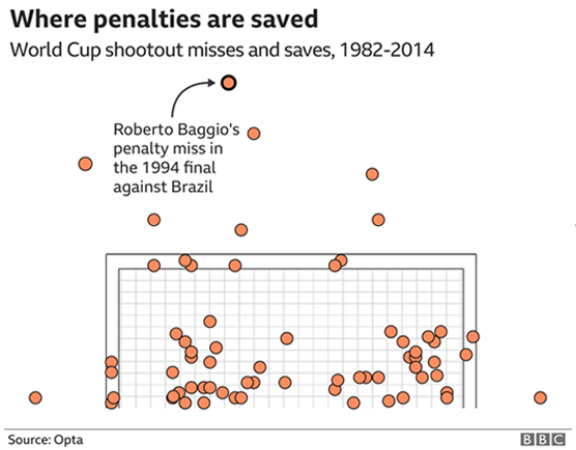
# Data Analysis in a Snap

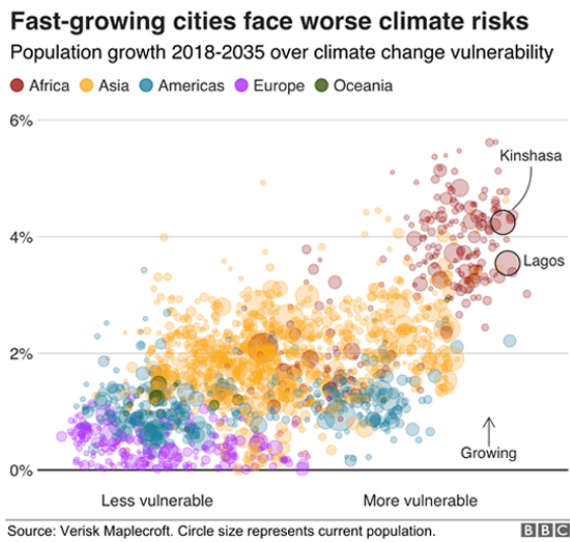
```
nhanes %>%
  group_by(gender) %>%
  drop_na(height) %>%
  summarize(mean_height = mean(height))
```

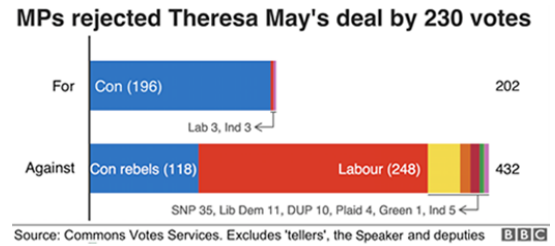
| gender | mean_height |
|---|---|
| female | 156.6159 |
| male | 167.1913 |

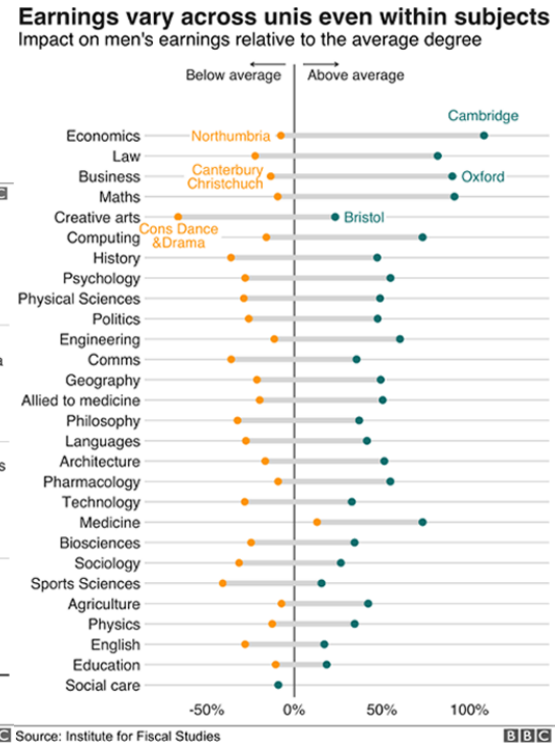
# High-Quality Data Visualization

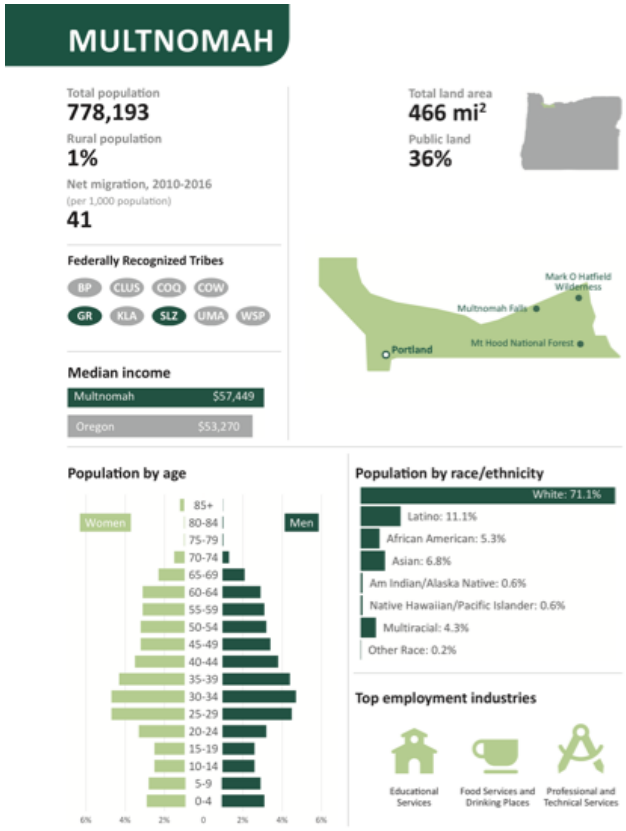

Blue wave
Won seat
Didn't win
Democrat candidates
Sharice Davids, the joint-first Native American woman in congress, increased Dem vote share by 13%
In West Virginia 3, Dems failed to turn the seat despite a 20% swing their way
30
20
10
0
Republican candidates
Republicans held onto Oregon 2 and Ohio 12, despite a 15% swing against them
30
20
10
0
–10%
0% change in vote share
+10%
+20%
Source: AP, 19:01 ET
BBC


Where penalties are saved
World Cup shootout misses and saves, 1982-2014
Roberto Baggio's penalty miss in the 1994 final against Brazil
Source: Opta
BBC


Fast-growing cities face worse climate risks
Population growth 2018-2035 over climate change vulnerability
Africa
Asia
Americas
Europe
Oceania
6%
4%
2%
0%
Kinshasa
Lagos
Growing
Less vulnerable
More vulnerable
Source: Verisk Maplecroft. Circle size represents current population.
BBC


MPs rejected Theresa May's deal by 230 votes
For
Con (196)
202
Lab 3, Ind 3
Against
Con rebels (118)
Labour (248)
432
SNP 35, Lib Dem 11, DUP 10, Plaid 4, Green 1, Ind 5
Source: Commons Votes Services. Excludes 'tellers', the Speaker and deputies
BBC


Earnings vary across unis even within subjects
Impact on men's earnings relative to the average degree
Below average
Above average
Economics
Law
Business
Maths
Creative arts
Computing
History
Psychology
Physical Sciences
Politics
Engineering
Comms
Geography
Allied to medicine
Philosophy
Languages
Architecture
Pharmacology
Technology
Medicine
Biosciences
Sociology
Sports Sciences
Agriculture
Physics
English
Education
Social care
Northumbria
Canterbury Christchuch
Cons Dance &Drama
Cambridge
Oxford
Bristol
-50%
0%
50%
100%
Source: Institute for Fiscal Studies
BBC

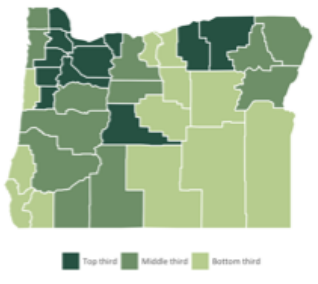





## MEDIAN INCOME

***Definition: The household income value at which 50% of households in the county earn less and 50% earn more.***

Median household income in this report provides a measure of the typical or "middle" income level in a county as well as the overall economic wellbeing for residents. One key drawback is that this measure treats all households equally regardless of the number of people in the household. The size of the household has a big impact on how the income is distributed to individuals. Nonetheless, median household income remains a broadly used measure. It is useful in tracking income growth, which is associated with the ability of residents to meet their needs, and comparing economic conditions across counties.




| Rank | County | Amount |
|---|---|---|
| 1 | Washington | $69,743 |
| 2 | Clackamas | $68,915 |
| 3 | Multnomah | $57,449 |
| 4 | Hood River | $56,581 |
| 5 | Columbia | $55,146 |
| 6 | Yamhill | $54,951 |
| 7 | Morrow | $54,441 |
| 8 | Deschutes | $54,211 |
| 9 | Polk | $54,010 |
| | *Oregon* | *$53,270* |
| 10 | Benton | $52,015 |
| 11 | Marion | $50,775 |
| 12 | Umatilla | $49,287 |
| 13 | Clatsop | $47,492 |
| 14 | Jefferson | $47,063 |
| 15 | Wasco | $46,814 |
| 16 | Linn | $46,782 |
| 17 | Jackson | $46,343 |
| 18 | Union | $45,564 |
| 19 | Lane | $45,222 |
| 20 | Tillamook | $43,777 |
| 21 | Wallowa | $42,349 |
| 22 | Douglas | $42,052 |
| 23 | Klamath | $41,951 |
| 24 | Baker | $41,722 |
| 25 | Sherman | $41,389 |
| 26 | Lincoln | $41,303 |
| 27 | Gilliam | $40,556 |
| 28 | Grant | $40,193 |
| 29 | Crook | $39,583 |
| 30 | Coos | $39,110 |
| 31 | Curry | $38,661 |
| 32 | Harney | $38,431 |
| 33 | Josephine | $37,867 |
| 34 | Malheur | $34,720 |
| 35 | Lake | $33,453 |
| 36 | Wheeler | $33,400 |

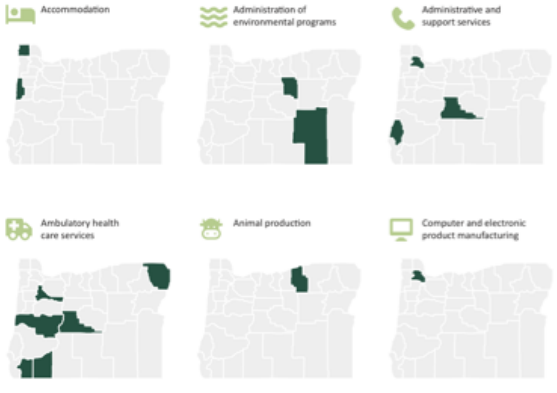
*Source: US Census Bureau, American Community Survey, Table B19013, 2012-2016, 5-year estimates updated annually. Released 2017.*



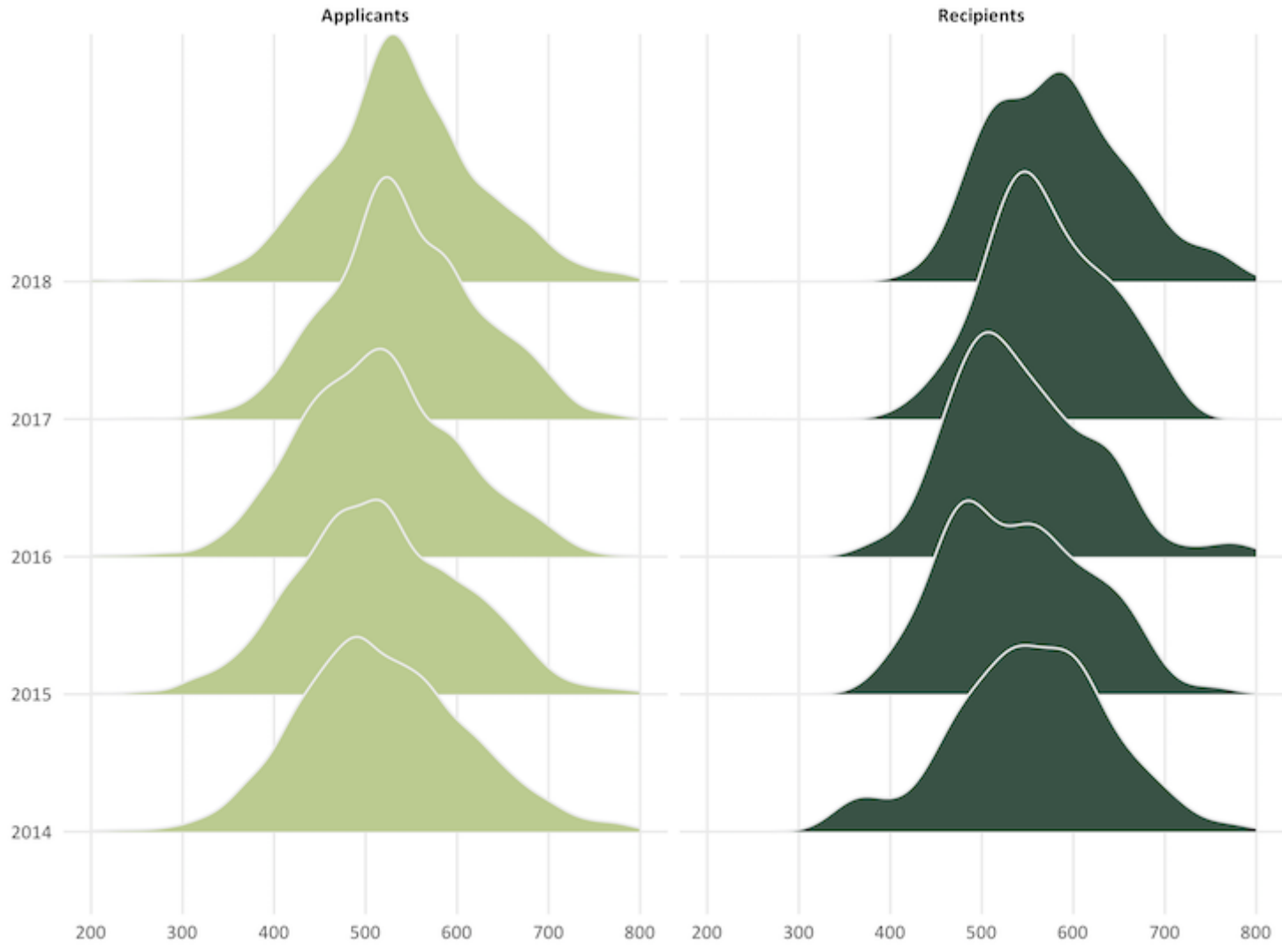
## TOP EMPLOYMENT INDUSTRIES

***Definition: The three industries with the greatest number of employees in each county, using the 3-digit North American Industry Classification System (NAICS) codes.***

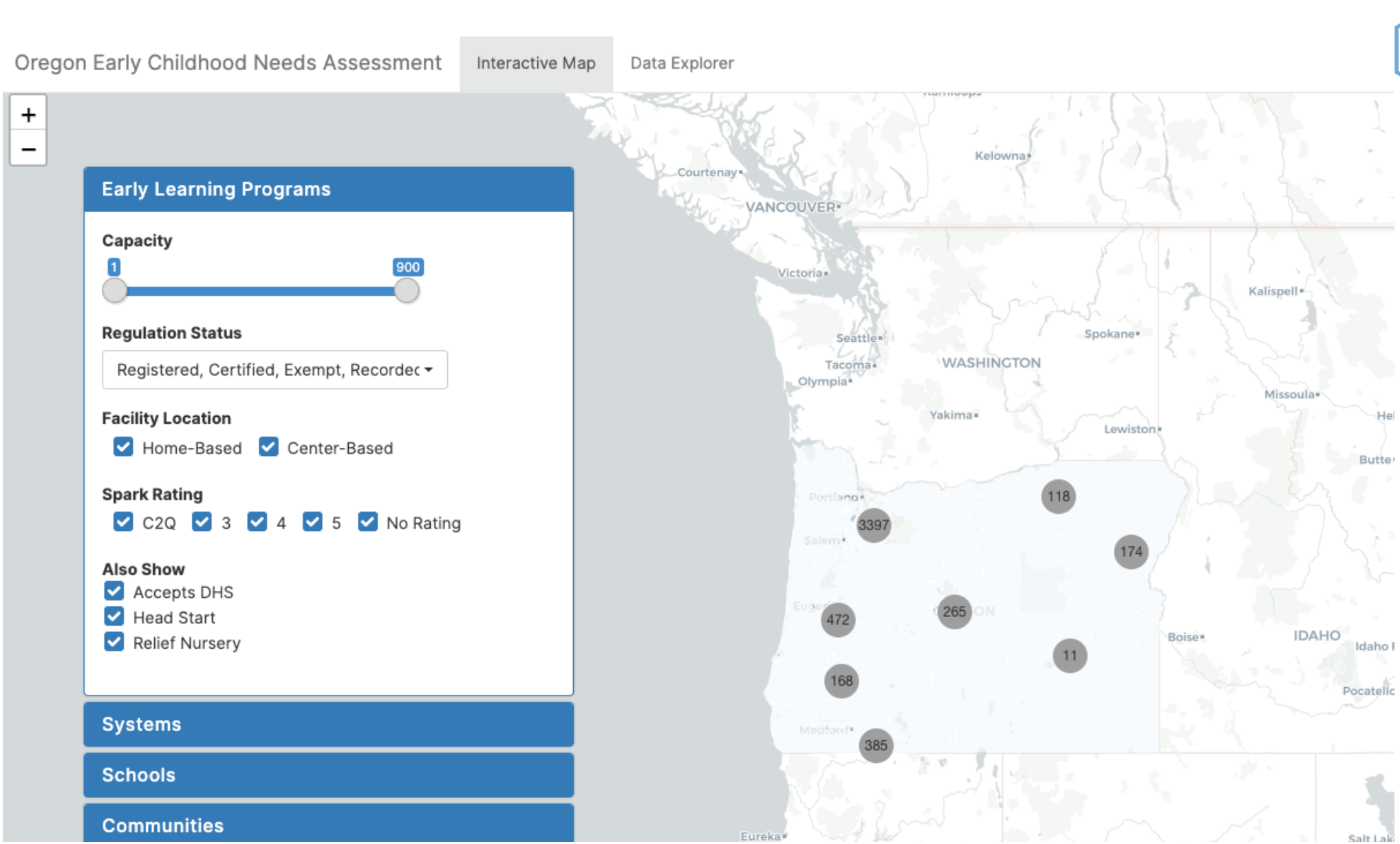
Identifying the top three employment industries in each county provides insight about the structure of the local economy. Employment industries have different average wage levels, so the top three figure prominently in determining the total wage earnings of a county. Examining this indicator across the state and between counties suggests notable employment trends and could point to policy opportunities. (Note: Each county profile shows the top three employment industries in ranked order from left to right.)



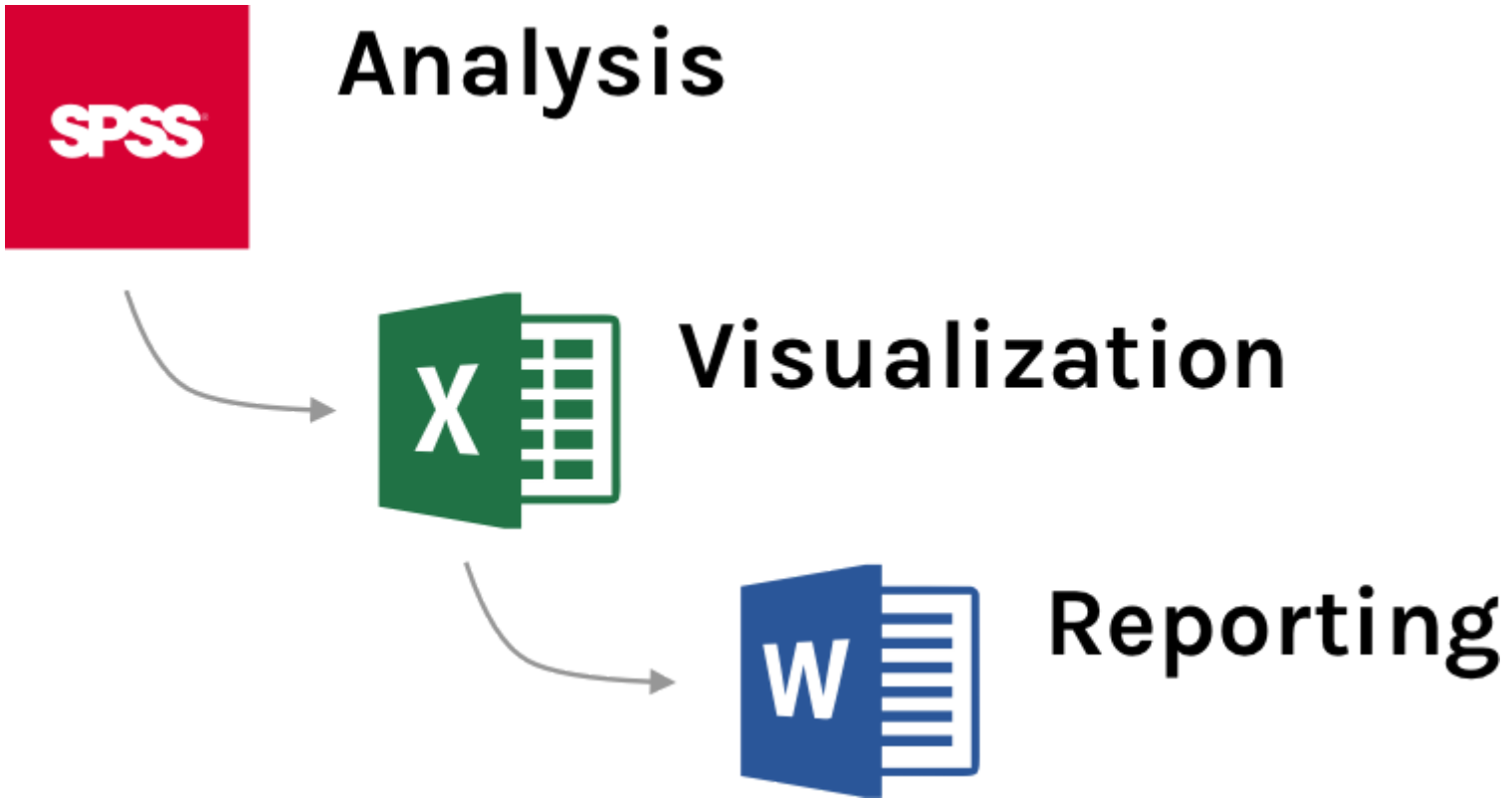

*Source: State of Oregon Employment Department, Economic Data 2016, updated annually. Released 2017.*


Applicants
Recipients
2018
2017
2016
2015
2014
200
300
400
500
600
700
800

# Unique Reporting Possibilities


Oregon Early Childhood Needs Assessment
Interactive Map
Data Explorer
+
−
Early Learning Programs
Capacity
1
900
Regulation Status
Registered, Certified, Exempt, Recordec
Facility Location
Home-Based
Center-Based
Spark Rating
C2Q
3
4
5
No Rating
Also Show
Accepts DHS
Head Start
Relief Nursery
Systems
Schools
Communities
Courtenay
Kelowna
VANCOUVER
Victoria
Seattle
Tacoma
Olympia
WASHINGTON
Spokane
Kalispell
Missoula
Yakima
Lewiston
Portland
Salem
Boise
IDAHO
Pocatello
Medford
Eureka
118
3397
174
265
472
11
168
385

# R's Killer Feature: RMarkdown


SPSS
Analysis
Visualization
Reporting

# RMarkdown

```
# Population

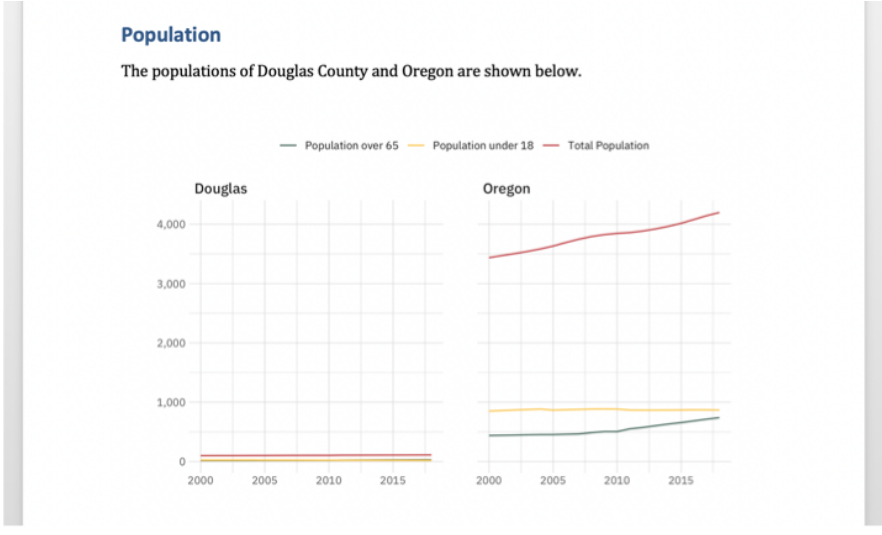
The populations of Douglas County and Oregon are shown below.

```{r}
population <- read_excel("data/dc-data.xlsx",
                         sheet = "Population") %>%
  clean_names() %>%
  gather("geography", "number", -c(indicator, year)) %>%
  mutate(number = number / 1000) %>%
  dk_replace_dc() %>%
  mutate(group = paste(indicator, geography))

ggplot(population, aes(year, number,
                       group = indicator,
                       color = indicator)) +
  geom_line() +
  facet_wrap(~geography) +
  scale_y_continuous(labels = comma_format()) +
  dk_remove_color_title +
  dk_set_colors


```
```

RMarkdown




Word

*[A]ll the work is done up front and then for every session …* ***I only need to spend 15 minutes generating the report and sending it to them****.*

Using R for Immediate Reporting in Evaluation by Dana Wanzer

# R Familiarity Survey

# The Best Reason to Learn R

# Questions?

# Start Your R Journey

# R in 3 Months (rfor.us/3months)

david@rfortherestofus.com

dgkeyes

rfortherest