Data Challenge

Department of Workforce Services

Workforce Research and Analysis

WORKFORCE SERVICES
RESEARCH & ANALYSIS

Data Challenge

- Develop an algorithm (or use existing algorithms) to identify the set of ***n*** Utah Census Tracts in ***k*** contiguous clusters (having a **minimum** cluster unemployment rate of 6.5 percent) that **maximizes** the total population across all ***n*** Census Tracts.

Why?

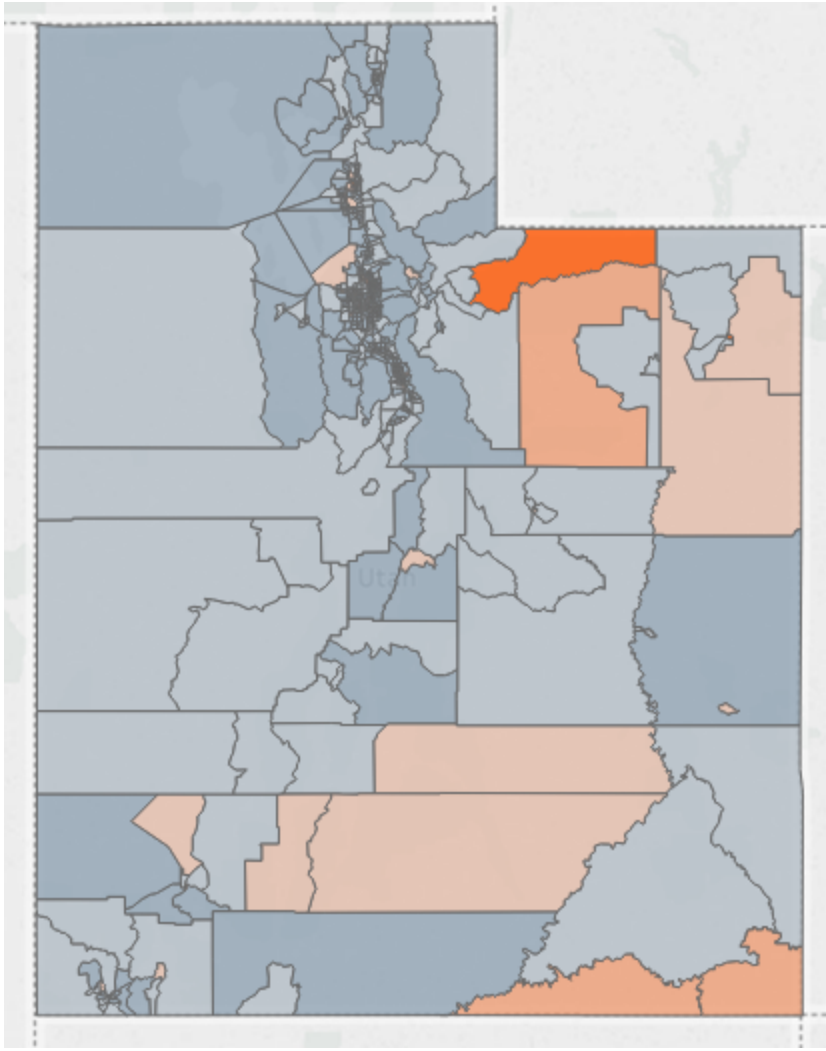
- The federal government has special funds to help local areas with substantial unemployment (ASUs).
- An ASU is defined as a cluster of Census Tracts with a cluster unemployment rate of **at least 6.5 percent**.
- Federal funding for programs to help these areas is distributed to the states based on the **total population of the ASUs** in the state.

Why?

- So – the state of Utah would like to identify the set of ASUs with maximum total population in order to maximize federal funding.

What the state currently does: Trial and Error

- We start with a map of all of Utah's Census Tracts.
- Each tract has three data points associated with it:
  - employment (***e***),
  - unemployment (***u***),
  - and population (***p***).



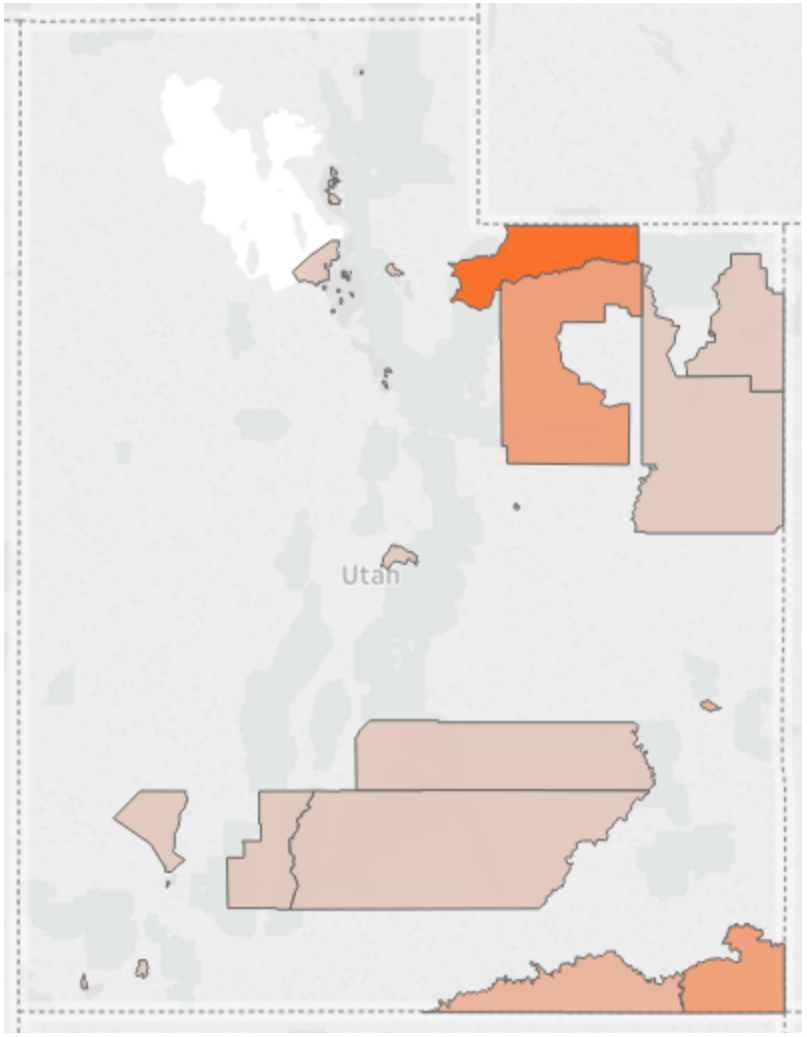
What the state currently does: Trial and Error

- We calculate the unemployment rate (***ur***) for each tract as follows:

$$u \,/\, (e \,+\, u) \,=\, ur$$

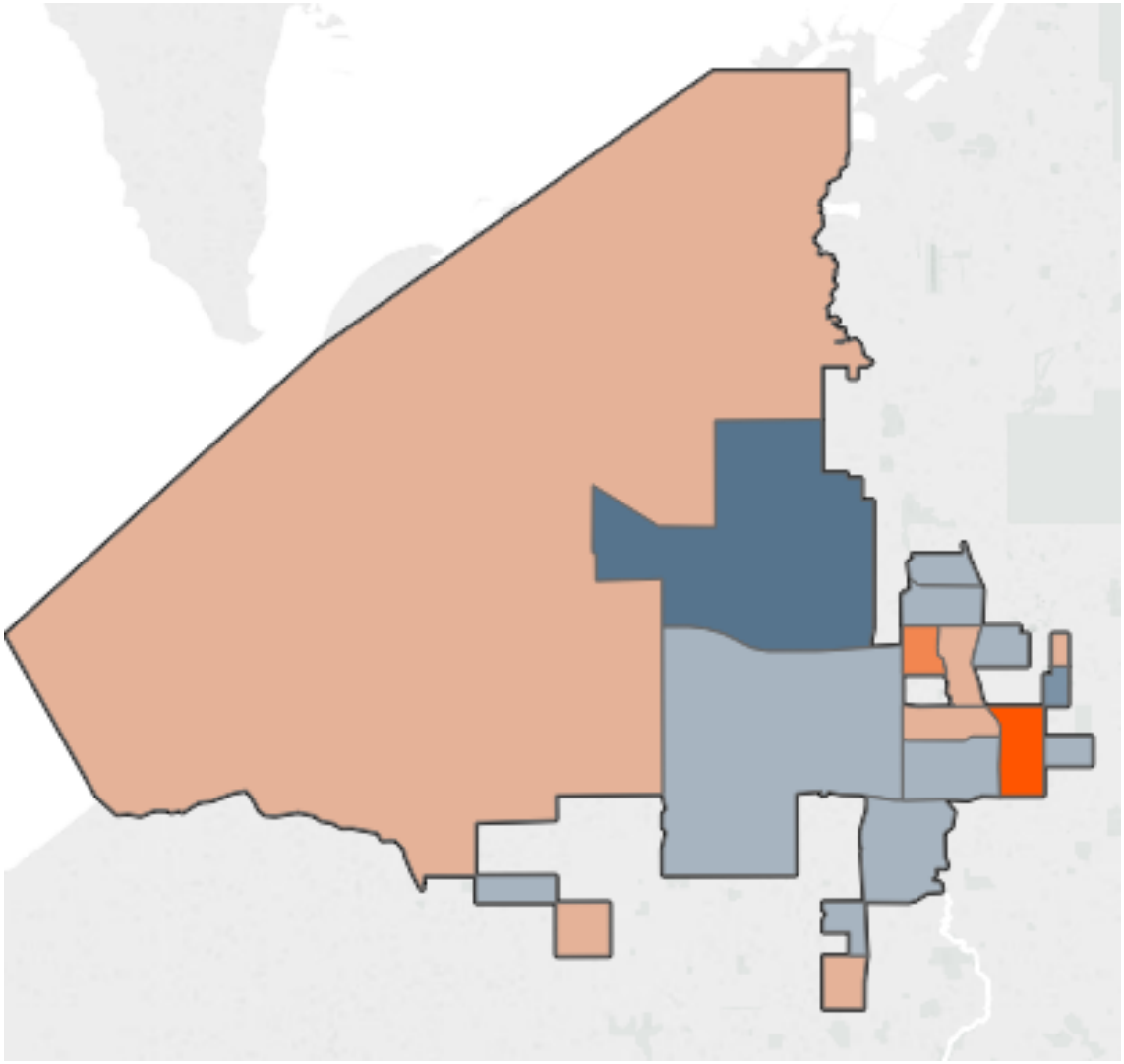
What the state currently does: Trial and Error

- We then identify which tracts - on their own – meet the $ur \geq 0.065$ criteria to be an ASU.
- All of these tracts will be included in a maximal solution.
- They might be what you would consider your "*seed*" tracts.



What the state currently does: Trial and Error

- Once we have those mapped, we can expand out a contiguous cluster (using queen contiguity) around each seed tract.
- The unemployment rate and total population must be re-calculated with each addition to make sure the cluster does not drop below 6.5 percent.
- If it does we remove that tract (or a different one if it results in higher total population).



Aggregate URate

6.53%

Aggregate 2015 Pop

91,202

What the state currently does: Trial and Error

Note:

- You can't just sum the unemployment rates as you add tracts to each cluster. You have to sum the components (i.e. ***u*** and ***e***) and then recalculate the ***ur***:

$$ur = \frac{\sum u_i}{\sum (u_i + e_i)}$$

What the state currently does: Trial and Error

- We continue this process by trial and error until we come up with a set of clusters – *each with an unemployment rate above 6.5 percent* – that has the maximum total population we can find.
- This is obviously very time consuming, and does not necessarily guarantee that we find the actual maximum possible outcome. **In fact it's very likely that we don't.**

The Data

The following files will be sent in an email:

- **Utah Census Tract CSV File:**

  UT_asu_exampleData.csv

  – This file contains 588 records (Census tracts) each with a unique identifier (“geoid”), and data fields for population (pop), employment (emp), and unemployment (unemp) .

- **2016 TIGER/Line Census Tract Shape File for Utah:** *tl_2016_49_tract.shp*

  – This file contains the polygon paths for all of Utah’s Census tracts (for determining spatial contiguity).

  – This file can be joined to the csv data on “Geoid”.

Resources that may be helpful

Tableau Calculator:

- https://public.tableau.com/views/ASU_Calculator_example/Dashboard1?:embed=y&:display_count=yes

Research Article and Pseudo Code:

"*Redistricting Using Constrained Polygonal Clustering*" – (Joshi et al., 2012)

- http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1107&context=csearticles

Python Library of Spatially Constrained Clustering Algorithms:

- clusterPy : http://www.rise-group.org/risem/clusterpy/

Submitting your answer

Include:

- your solution (i.e. the list of census tracts)
- your code
- an explanation of how it works
- and any info that we might need to run it.