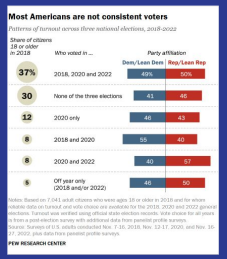
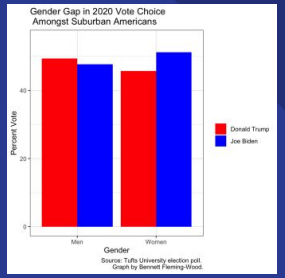
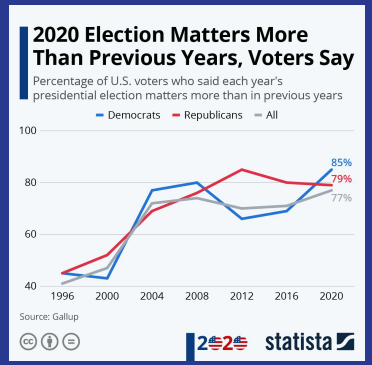
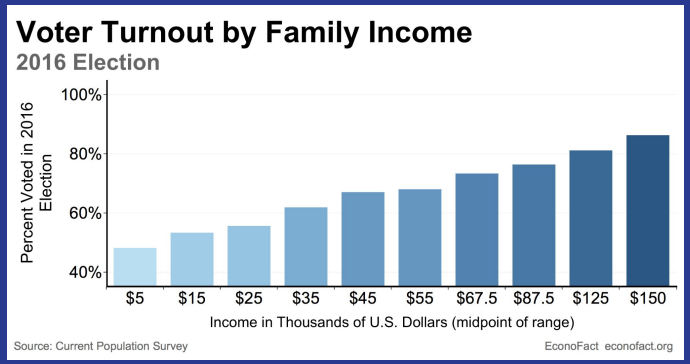




Elections, Education, Unemployment Data 351 Final Project

AJ Tennathur, Jerrick Little





The Datasets - Brief Overview

Elections

- Voting information/count/percentages by County for ***every*** county in America
- Education levels
- Goals of our analysis: **Trends based on various factors included/not included within dataset**
- What aspects of this dataset did we actually utilize?

Unemployment

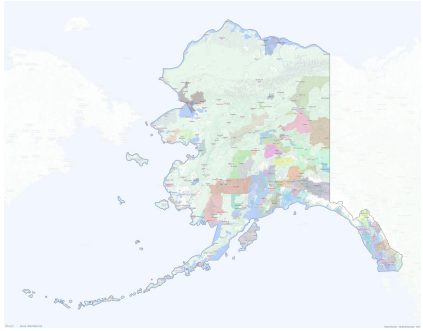
- Unemployment rate of ***adults*** in the USA by ***every*** county
- Year: 2020, match the election analysis dataset
- What aspects of this dataset did we actually utilize vs all the data provided?

FIPS Codes

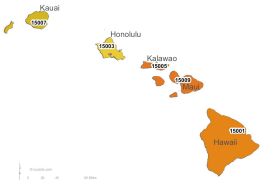
FIPS Codes: Numeric codes that uniquely identify geographic areas in the United States

- Dataset contains the FIPS codes for ***every*** county in America

Data Cleaning/Key Creation Process:
Unemployment Data

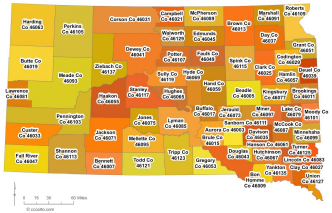


- ★ Note: **Original Data Size** – not just 2020 data
- ★ NA removal
- ★ Puerto Rico, United States Cumulative Data
- ★ FIPS Codes removal
  - 02201→ Prince of Wales - Outer Ketchikan Census Area
    - **No longer exists**, 2198
  - 02232 → Skagway-Hoonah-Angoon Census Area
    - **Split,** 2 FIPS Codes, 2230 and 2105
  - 02280 → Wrangell-Petersburg Census Area
    - **Split**, 2 FIPS Codes, 2195 and 2275
- ★ **Keys**
  - FIPS_Code column → primary key, references FIPS



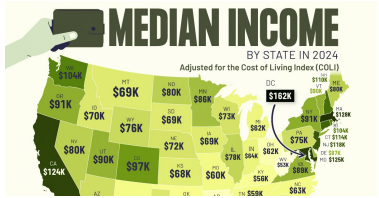
Data Cleaning/Key Creation Process:
FIPS Data

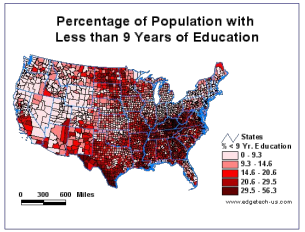
- ★ United States, Puerto Rico - removed
- ★ FIPS Codes Removals
  - **15005** → Kalawao County, HI → removed by us
  - **51515** → Bedford City, VA → **merged** with Bedford County, FIPS **51019**
- ★ Renaming County Names
  - FIPS Code **46113** – Shannon County, SD
    - Changed to Oglala Lakota County, new FIPS **46102**
  - FIPS Code **02270** – Wade Hampton Census Area, AK
    - Changed to Kusilvak Census Area, new FIPS **02158**

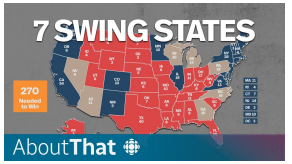


- ★ **Keys**: Fips column → primary key, references FIPS in unemployment dataset
- ★ Note: We did not ADD data. The data for the "new" FIPS codes was already in the data, we are taking out data that is outdated and no longer used or updated.

Data Cleaning/Key Creation Process:
Election Data

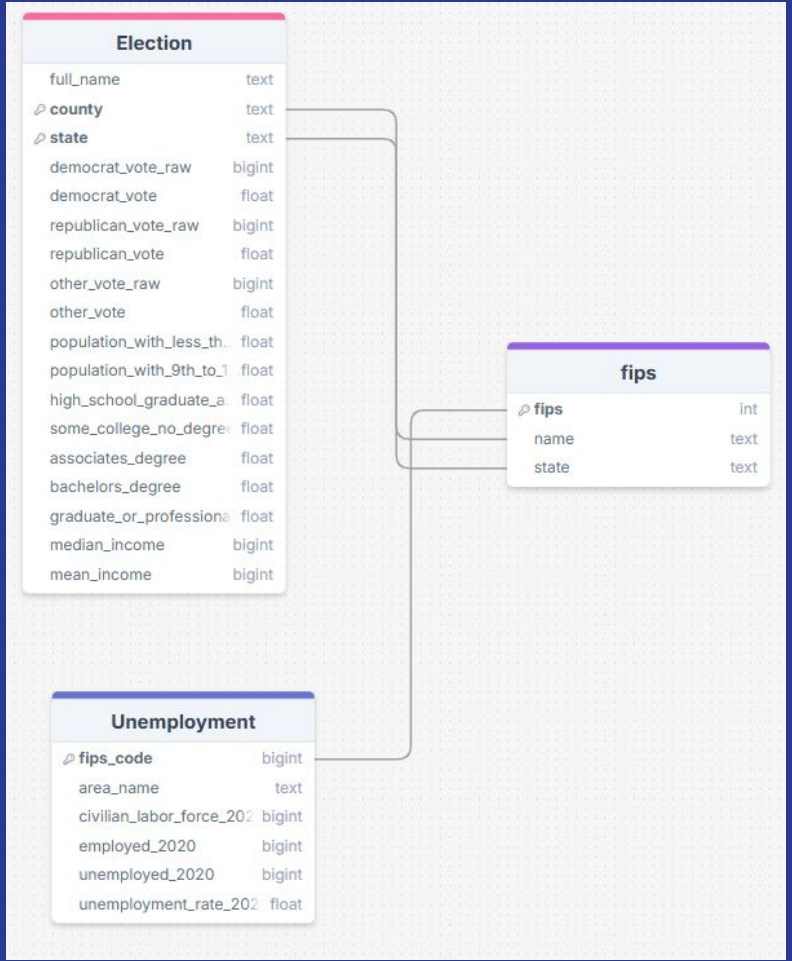






- ★ Note: **Original Data Size**
  - Race, ethnicity, immigration, job sectors, Gini Index, & more!
- ★ What we chose to keep:
  - **Voting by year (2020), county, state, education levels , income levels**
- ★ **Keys:** county, state as a composite primary key
  - primary key (county, state)
  - There are counties with same names across states (Nassau County)
- ★ Ex: WHEN state = 'Arizona' THEN 'AZ' → keep data uniform and easier to join

Simple ERD and Tables



Constraints

```
alter table unemployment
add constraint unemp_fips_key primary key (FIPS_Code);


alter table unemployment
add constraint unemp_fips_fkey foreign key (FIPS_Code)
references fips (fips);


alter table fips
add constraint fipstable_fips_key primary key (fips);

alter table fips
add constraint fipskey_fips_fkey foreign key (fips)
references unemployment (FIPS_Code);

ALTER TABLE unemployment
ADD CONSTRAINT chk_labor_force
CHECK (Civilian_labor_force_2020 = Employed_2020 + Unemployed_2020);

ALTER TABLE election
ADD CONSTRAINT chk_total_vote
CHECK (democrat_vote + republican_vote + other_vote = 100);
```

What do we want to know?

How does unemployment and education affect election results?

Question 1:

- What is the **average unemployment rate** in counties where the percentage of people with only **high school diplomas** is **between 20 and 30 percent and voted democrat?**
- Answer: 8.32%

Question 1 Code

```
SELECT avg(u.unemployment_rate_2020)

FROM unemployment as u

JOIN fips as f ON f.fips = u.fips_code

JOIN election as e ON e.county = f.name AND e.state = f.state

WHERE e.high_school_graduate_and_equivalent BETWEEN 20 AND 30

AND e.democrat_vote > e.republican_vote;
```

Learning Objectives

- I can compute many different aggregate calculations over an entire column.
- I can use a basic inner JOIN to combine information from several tables appropriately
- I can write a SELECT query that utilizes basic WHERE filtering and ORDER BY

Question 2:

- What percentage of counties with at least 10% of the population holding **graduate degrees**, an average income of **$70,000** or more,more than **100,000 employed** individuals, and **voted republican**, also have **unemployment rates below 5%**?
- Answer: .69% of counties nationwide

Learning Objectives

- I can use CTE's to precompute information for later use within a query.
- I can use a basic inner JOIN to combine information from several tables appropriately
- I can write a SELECT query that utilizes basic WHERE filtering and ORDER BY
- I can properly use DISTINCT in a query to answer a question.

Question 2 Code

```
WITH filt_counties AS (SELECT COUNT (DISTINCT e.county) as county_count FROM election as e

JOIN fips as f ON f.name = e.county AND f.state = e.state

JOIN unemployment as u ON u.fips_code = f.fips

WHERE e.graduate_or_professional_degree >= 10 AND e.mean_income >= 70000 AND u.employed_2020 > 100000 AND
u.unemployment_rate_2020 < 5 AND e.republican_vote > e.democrat_vote),

total_counties AS (

  SELECT COUNT (DISTINCT county) as total_count

  FROM election)

  SELECT (filt_counties.county_count::FLOAT / total_counties.total_count) * 100 AS percentage FROM filt_counties,
total_counties;
```

Questions?