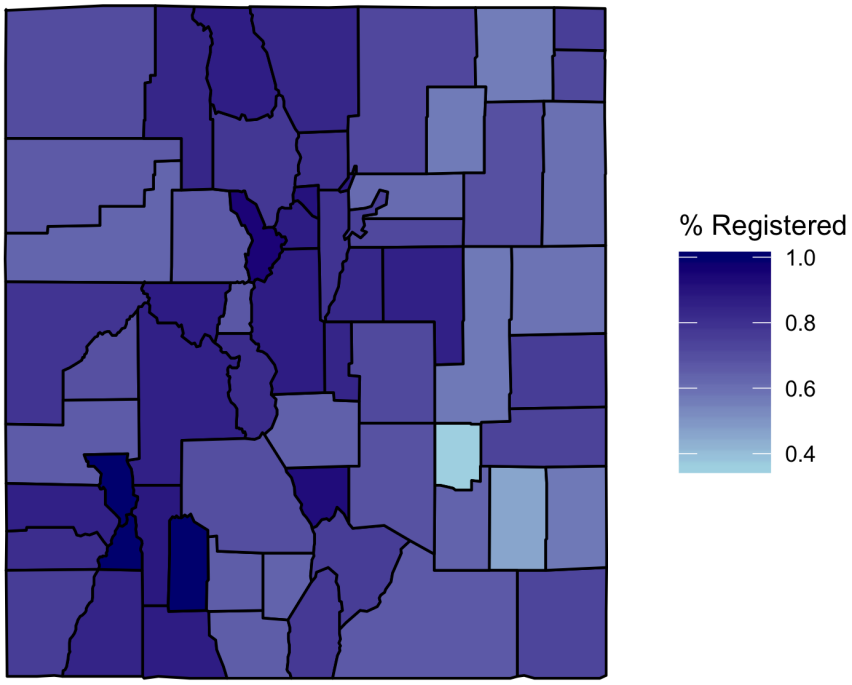

Figure 3.2: Registration rates per Colorado county

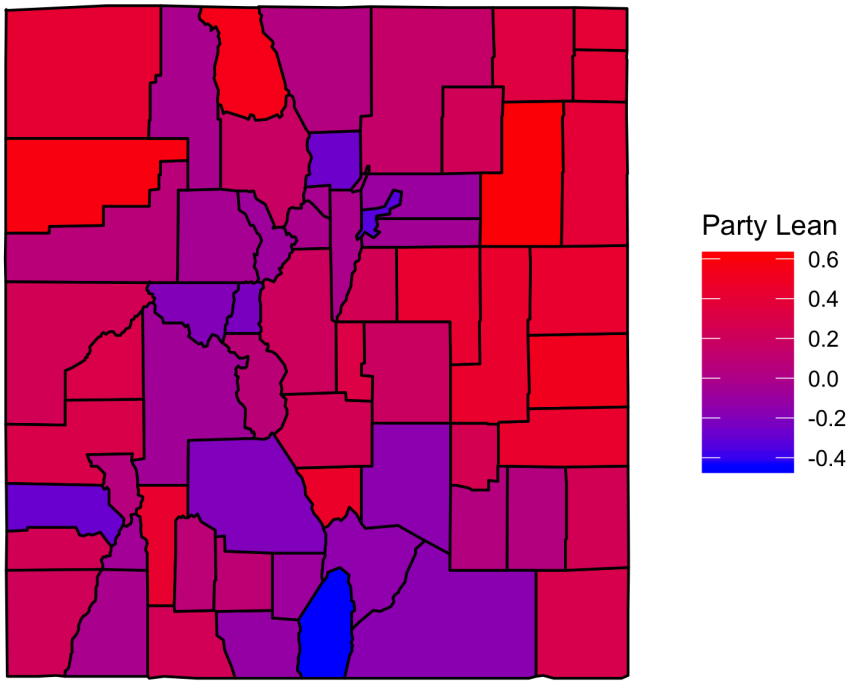

Figure 3.3: Democratic/Republican party lean per Colorado county

Table 3.3: Key changes to Colorado elections policy

Year	Key Changes
1992	No Excuse Absentee Statewide Implementation
2008	Permanent No-Excuse VBM Lists, Option of Full-VBM Elections
2013	Automatic Mail Ballot System Implemented Statewide, Established Vote Centers

3.1.4 Colorado as a Case for this Thesis

Colorado presents an interesting case for research on Vote By Mail exactly because it has gone through a long transitional process to reach its current elections system. It has steadily developed voting policy through a mixture of state mandates, county action, and outside policy motivations. Colorado's streak of independence and direct democracy is also very apparent in this shift in electoral practices, since they have been passing policies trying to expand participation for a very long time. It gives researchers access to approximately 22 years during which at least part of the state conducted elections by mail, making comparative, county- or individual-level case studies particularly alluring. Colorado's streak of independence and direct democracy is also very apparent in this shift in electoral practices, since they have been passing policies trying to expand participation for a very long time.

On a more general level, Colorado is interesting exactly because it is "typical" but with a wild streak. It is typical rocky mountain country, great plains country, and liberal urban city but all *in one state*. In is libertarian yet increasingly Democratic. It heavily relies on state funding for national parks, yet rebels against federal land use laws. Colorado overwhelmingly supports marijuana legalization, despite being a frontier state with traditional values. It is also a consistent purple state, with a Democratic Governor and House, but Republican Attorney General, Secretary of State and Senate. This means that Colorado is a combination of distinct national effects, but also local effects that make it significantly different from national trends as a whole. In this environment, predicting results of policy can be difficult, but extremely salient as multiple effects can be tested against each other.

3.2 The Data

This thesis relies on county and individual level models to draw conclusions on voting behaviors, and how they are affected by voting method. As such, the data I need will optimally contain the following:

- **County and individual level demographic characteristics**: race, gender, urban population
- **County and individual level voting data**: turnout, party registration, total registrants

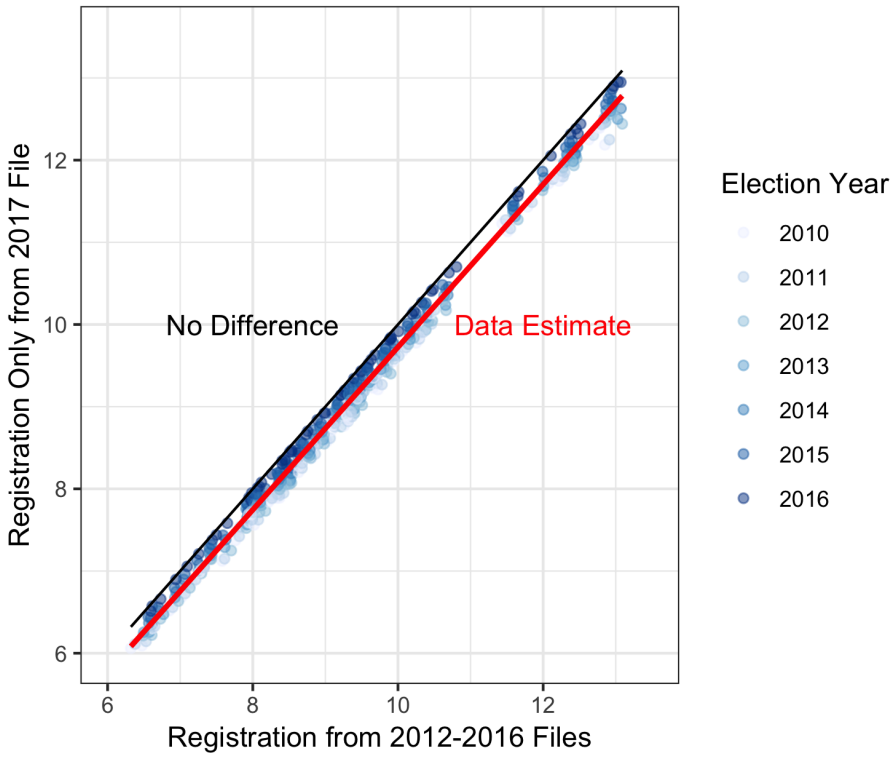

Figure 3.5: Comparison of registration count methods

My selection was later vindicated, when looking at comparisons between reported rates of turnout[5] and turnout calculated through my dataset for the 2014 midterm election (see fig. 3.7).

The differences are insignificant. They exist because of "noise" added on because of errors in the data, misreporting, registration records redacted due to privacy concerns, voters dropped before the "snapshot" occurred, and other similar factors.

3.3.2 Other Wrangling Issues

Wrangling the data was the majority of the work that went into this thesis. As will become clear in this section, apart from accurately processing, diagnosing, and merging the data, the process of wrangling includes several non-trivial decisions about how to treat missing values and variable specification. Including a full account would probably read like the world's most cliche crime novel: a series of elusive final datasets, a plucky yet occasionally naive young detective, two wisened mentors, clues, dead ends, frustration, compromise, and... spreadsheets. I will spare the reader the whole story, but I will include a non-comprehensive list of some of the difficulties associated with wrangling voter files, as it was a crucial part of the learning process I underwent while doing my research.

Missing Values: The decision on how to deal with missing values–or NAs–in a

[5]Turnout is calculated over all registered voters.

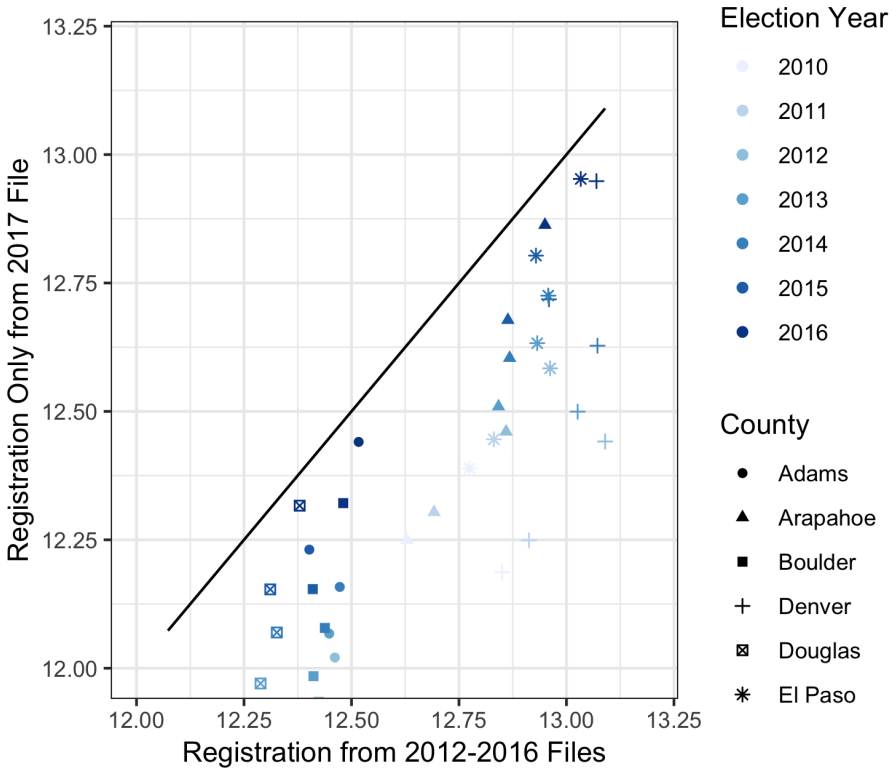

Figure 3.6: Comparison of registration count methods only for a few counties, 2012-2016