

# Environmental Statistics Final

Halina Do-Linh

12-02-2021

## Question

### Are we drilling deeper wells over time in California?

1.5 million people in California rely on domestic well water. (Johnson) Majority of the 1.5 million reside in rural areas without alternative drinking-water sources. Furthermore, in the Central Valley, it has been found that one in five wells now run dry due to groundwater level declines (Jasechko 1). While some have responded to this issue by drilling deeper, this practice has also been found to be unsustainable because it is costly, it is impractical in certain hydrogeological conditions, and deep groundwater is often brackish or saline. (Perrone 3) While we are aware that groundwater levels are declining and that deeper drilling is unsustainable, we do not know if this is a long-term trend in California.

It is important to study groundwater trends because water is an important resource in California and it is crucial for groundwater management to be sustainable and equitable.

## Data

My data has been obtained by Dr. Debra Perrone and Dr. Scott Jasechko for their research on groundwater wells. California domestic well water data was obtained through personal communication with state representatives from the California Department of Water Resources (CADWR) in March 2016 and April 2016. The CADWR maintains the California Groundwater Well Completion Report database. Dr. Perrone and Dr. Jasechko reported that the data may contain biases by region, time, purpose, and compliance.

This data is international, but I subsetted the data to only contain wells from California. I also added region codes based on the California Complete Count Office, which grouped California's 58 counties into 10 regions based on hard-to-count populations, like-mindedness of the counties, capacity of community-based organizations within the counties, and state Census staff workload capabilities.

## Data Histograms

After taking the natural log of well depth (m), it appears to have a normal distribution . It also looks like majority of wells are located in Region 1 or Superior California (Figure 1).

Exploratory graphs of total counts of wells both over time and by region suggests that overall we are completing less wells (figure 1). However, we are also seeing that we could be potentially drilling deeper wells over time (figure 2).

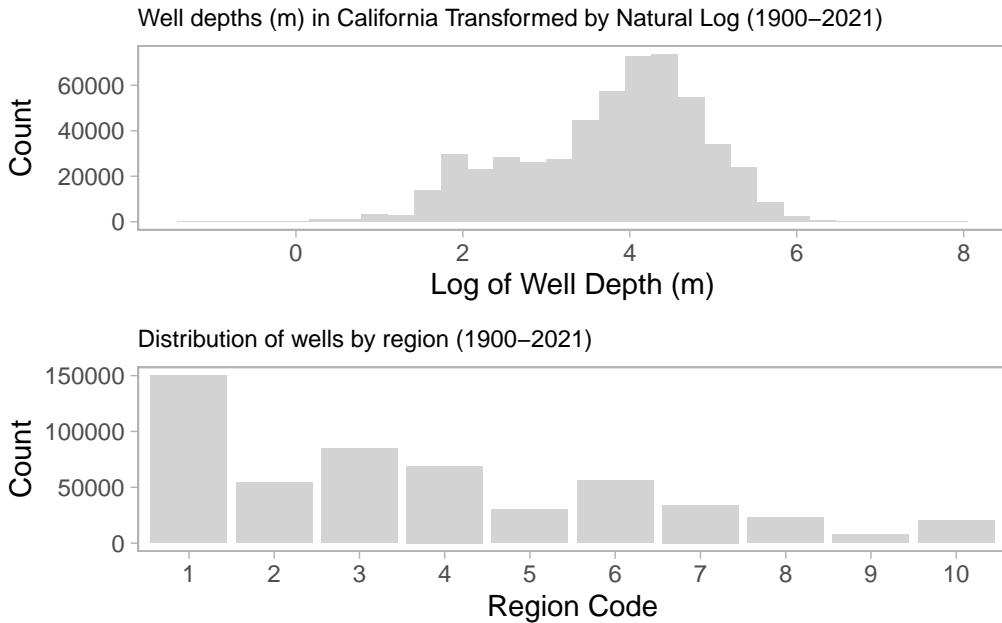


Figure 1: Well depth (m) data transformed by natural log were normally distributed (top). Majority of wells in California can be found in Northern regions of California identified here as region codes 1-5 (bottom).

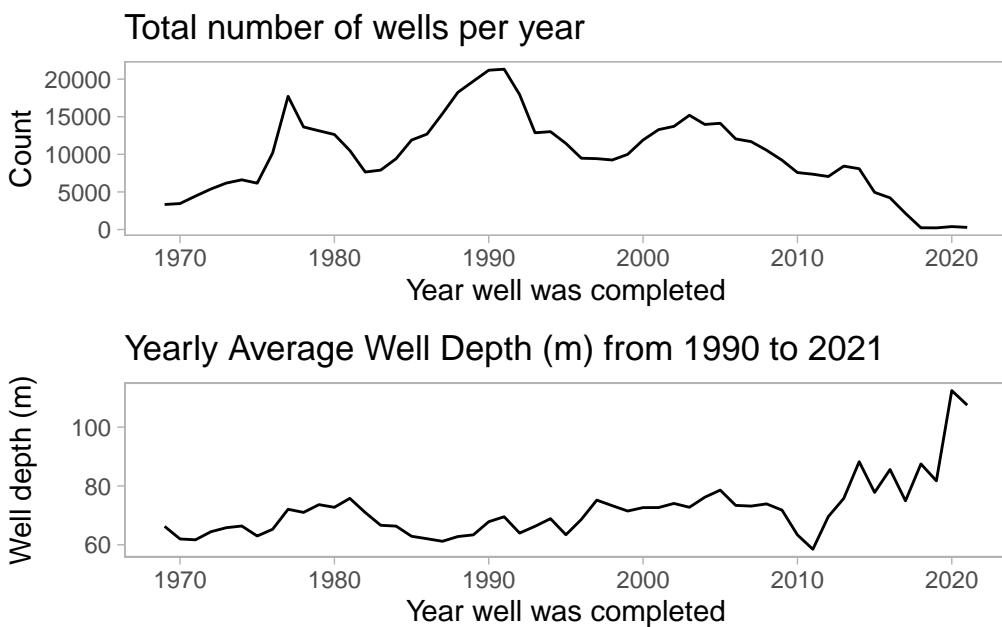


Figure 2: Number of wells are appearing to decrease overtime in California with a significant drop in 2005 (top). Average well depth (m) are appearing to increase overtime in California with a significant increase after 2010 (bottom).

# Analysis Plan

## 1. Regression Analysis

Use a multiple linear regression model to test if there is an effect of time and region on well depth (m).

$$\text{welldepth}(m) = \beta_0 + \beta_1 \cdot \text{year}_i + \beta_2 \cdot \text{region}_i + \varepsilon_i$$

I added an interaction effect to my multiple linear regression model to see if the effect of year on well depth (m) depends on region. I predict there may be deeper wells in the northern regions than the southern regions since majority of California's drinking water comes from northern regions. An interaction effect allows me to test my prediction.

$$\text{welldepth}(m) = \beta_0 + \beta_1 \cdot \text{year}_i + \beta_2 \cdot \text{region}_i + \beta_3 \cdot \text{year}_i \cdot \text{region}_i + \varepsilon_i$$

## 2. Time Series Analysis

Run decomposition analysis to test for seasonality. From my exploratory graphs I visually saw some evidence of cyclical patterns (see Figure 6 in Appendix) and wanted to see if this was significant.

# Results

## Multiple Linear Regression

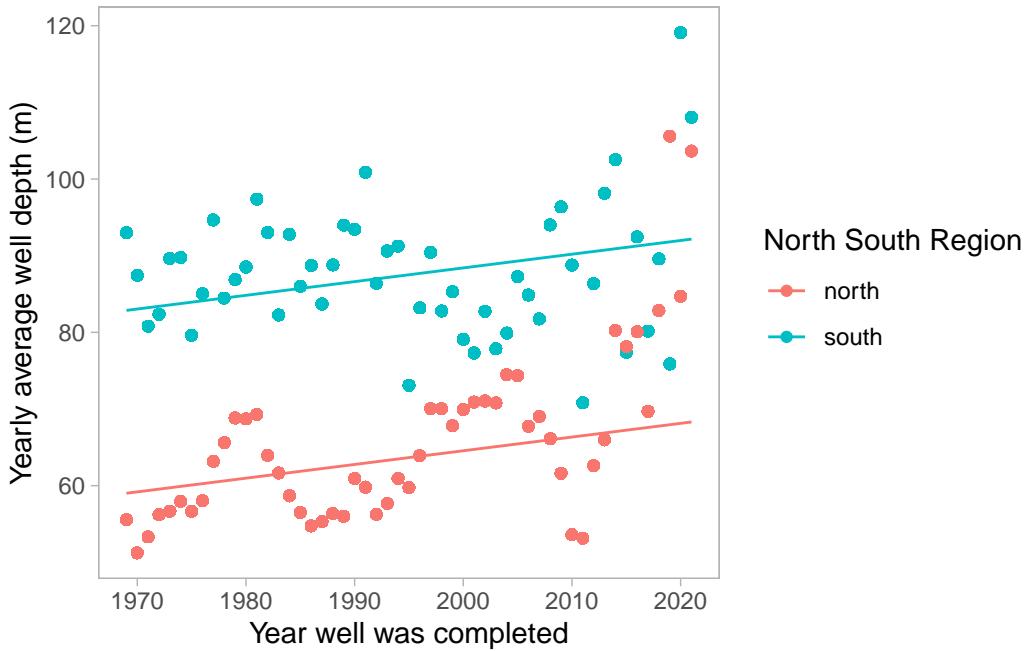


Figure 3: Multiple linear regression model of year and region on well depth. P-value < 0.0000000000000022. Adjusted R-value 0.7354648.

**Interpretation:** The intercept coefficient tells us that we expect to see a well depth of -292.98 m at year 0 in the south region (this is nonsensical). The `year_completed` coefficient tells us that we expect well depth to increase by 0.178 m each year, holding region constant. The `north_south` coefficient tells us that we

expect to see, on average, a 23.85 m increase in well depth in the south region compared to the north region, holding `year` constant.

This parallel slopes model tells us that `year`, regardless of `region`, will have the same impact on yearly average well depth (m).

## Mulitple Linear Regression Model with Interaction Effect

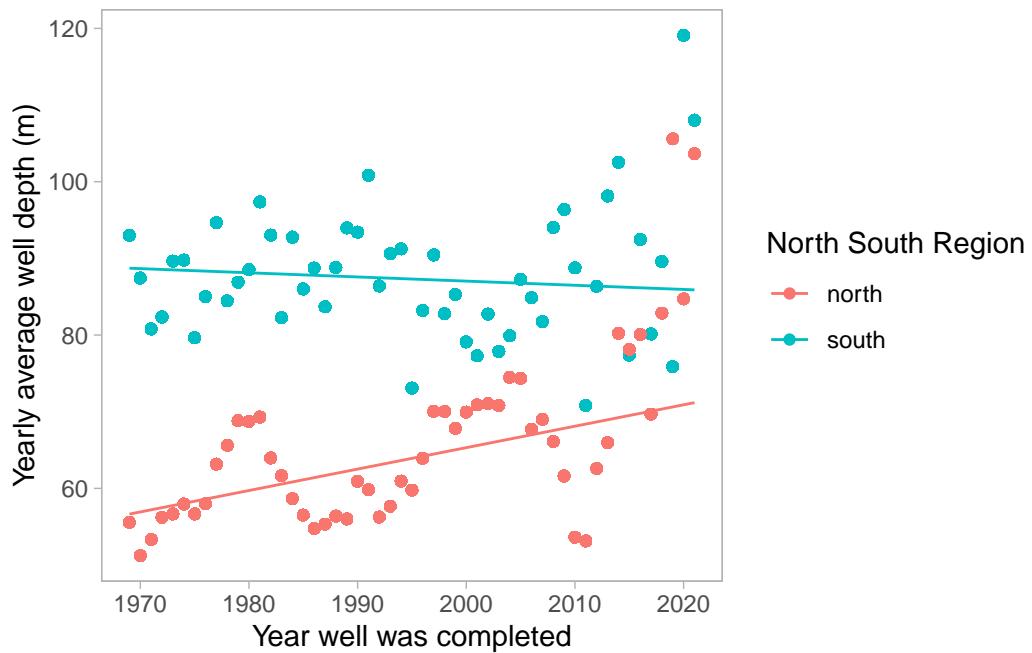


Figure 4: Interaction effect of year and region on well depth. P-value < 0.00000000000000022. Adjusted R-value 0.7563347

**Interpretation:** The `intercept` coefficient tells us that we expect to see a well depth of -493.84 m at year 0 in the north region (this is nonsensical). The `year_completed` coefficient tells us that we expect well depth to increase by 0.27 m each year for a well located in the northern region. The `north_south` coefficient tells us that well depths are, on average, 689.50 m deeper in the north region than the south region at year 0 (this is also nonsensical). The `year_completed:north_south` coefficient tells us that `year` on yearly average well depth (m) is 0.33 higher in the southern region than the northern region. The relationship between `year` and `well_depth` varies with `region`.

Additionally, since the adjusted  $R^2$  value increased slightly, this tells us that the interaction model improves the model fit a little. This indicates that `north_south` adding value to the model.

## Time Series Analysis - Decomposition

**Interpretation:** - `trend`: We see that `trend` is adding value to the model. - `seasonal`: We do not see that `seasonal` is adding value to the model. - `random`: We see that `random` is adding value to the model. - `mean_depth_m_year`: We see that `mean_depth_m_year` is adding value to the model.

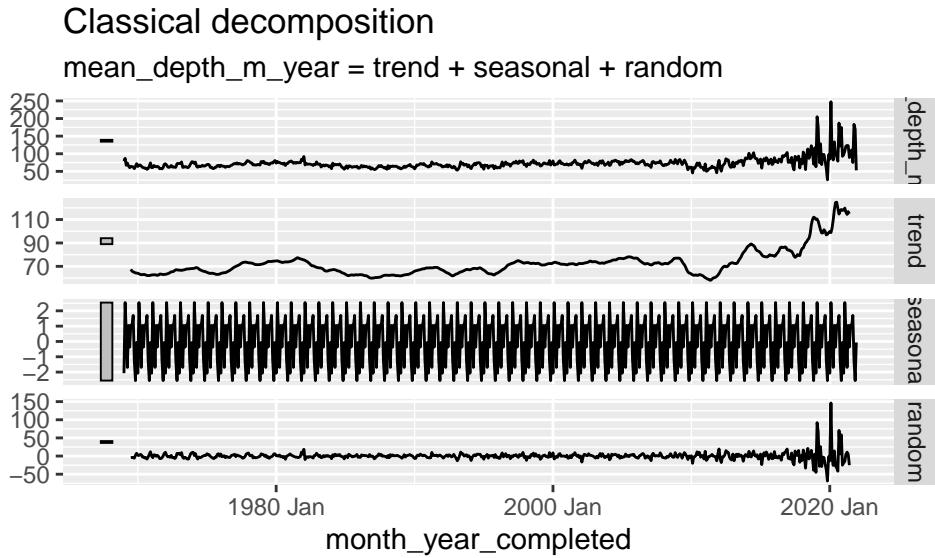


Figure 5: Decomposition of yearly mean well depth (m) with significance with trend, but not seasonal.

## Future Research

Future research would check to see if seasonality is present with just northern regions.

## Appendix

### Figures

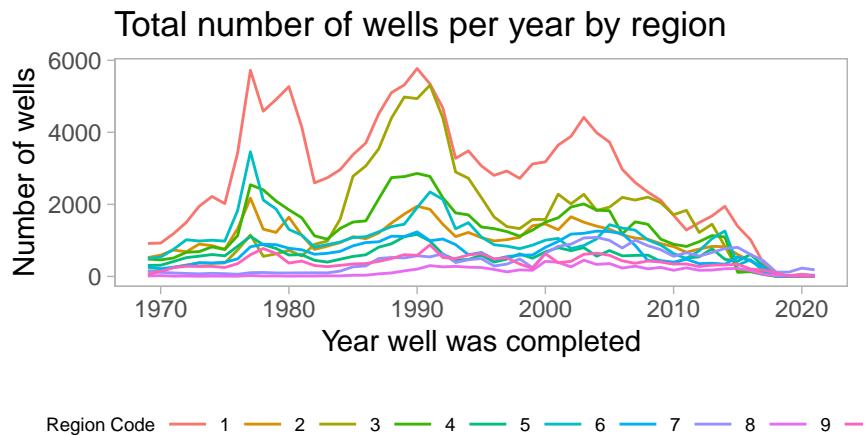


Figure 6: Total number of wells by region with potential for cyclical or seasonal trending.

## References

1. Jasechko, Scott, and Debra Perrone. "California's Central Valley Groundwater Wells Run Dry During Recent Drought." *Earth's Future*, vol. 8, no. 4, Apr. 2020. DOI.org (Crossref), <https://doi.org/10.1029/2019EF001339>.
2. Johnson, Tyler D., and Kenneth Belitz. "Identifying the Location and Population Served by Domestic Wells in California." *Journal of Hydrology: Regional Studies*, vol. 3, Mar. 2015, pp. 31–86. DOI.org (Crossref), <https://doi.org/10.1016/j.ejrh.2014.09.002>.
3. Perrone, Debra, and Scott Jasechko. "Deeper Well Drilling an Unsustainable Stopgap to Groundwater Depletion." *Nature Sustainability*, vol. 2, no. 8, Aug. 2019, pp. 773–82. DOI.org (Crossref), <https://doi.org/10.1038/s41893-019-0325-z>.