Let's Talk About the Weather

Jill Lundell, Brennan Bean, Jürgen Symanzik - Department of Mathematics and Statistics, Utah State University

Overview | Clusters | Seasonality | Trends & Outliers | Importance & Correlations | Conclusions

Motivation

How do different regions of the United States experience forecast error? (click on images or tabs for details)

Clusters

Cali-Florida

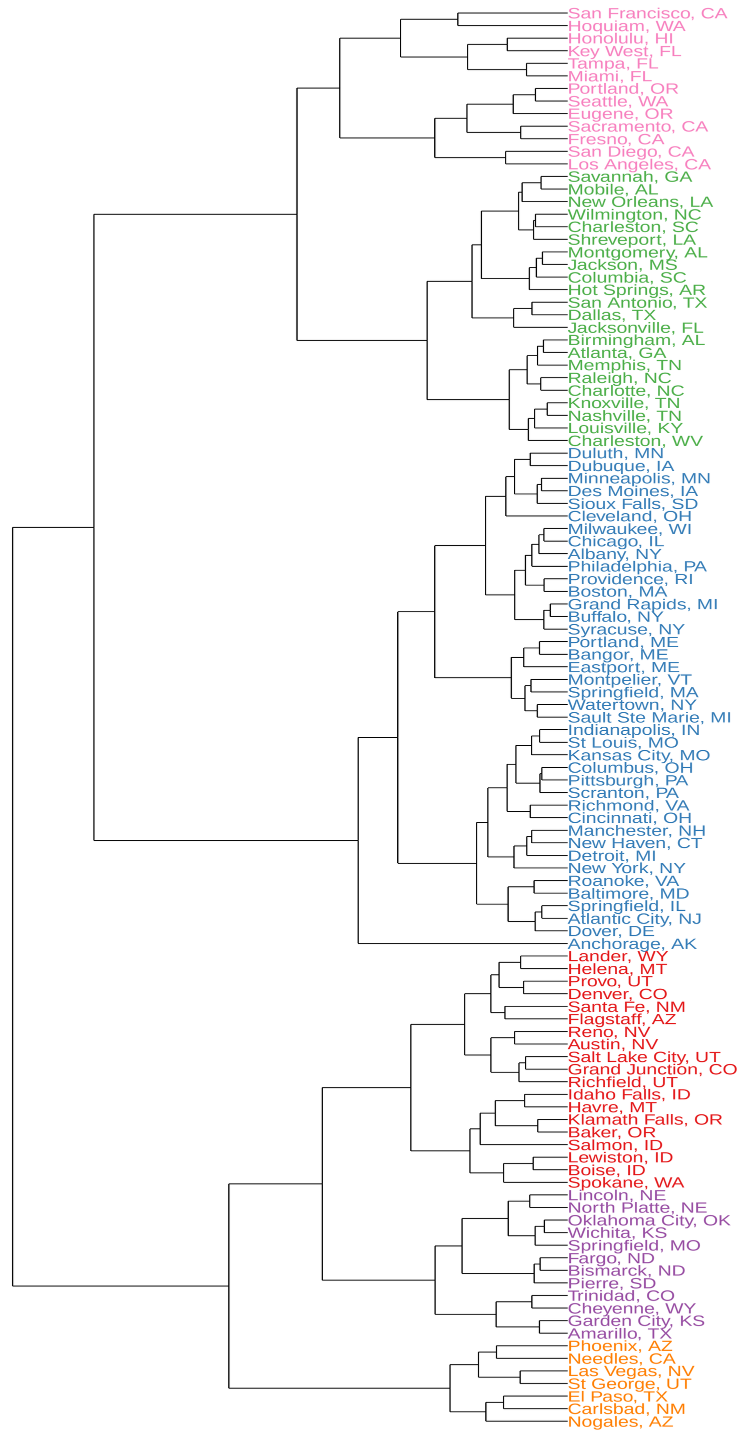
Intermountain West

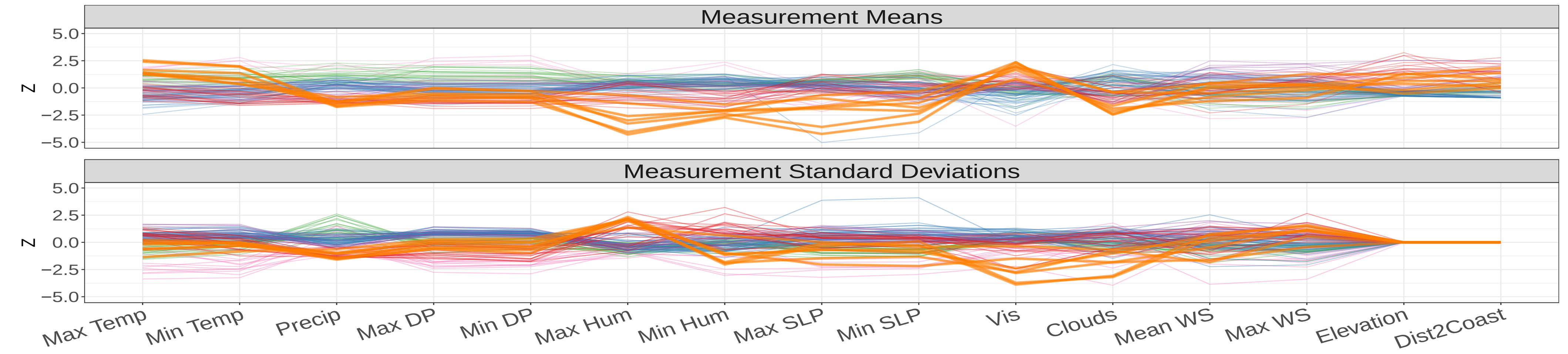
Southeast

Midwest

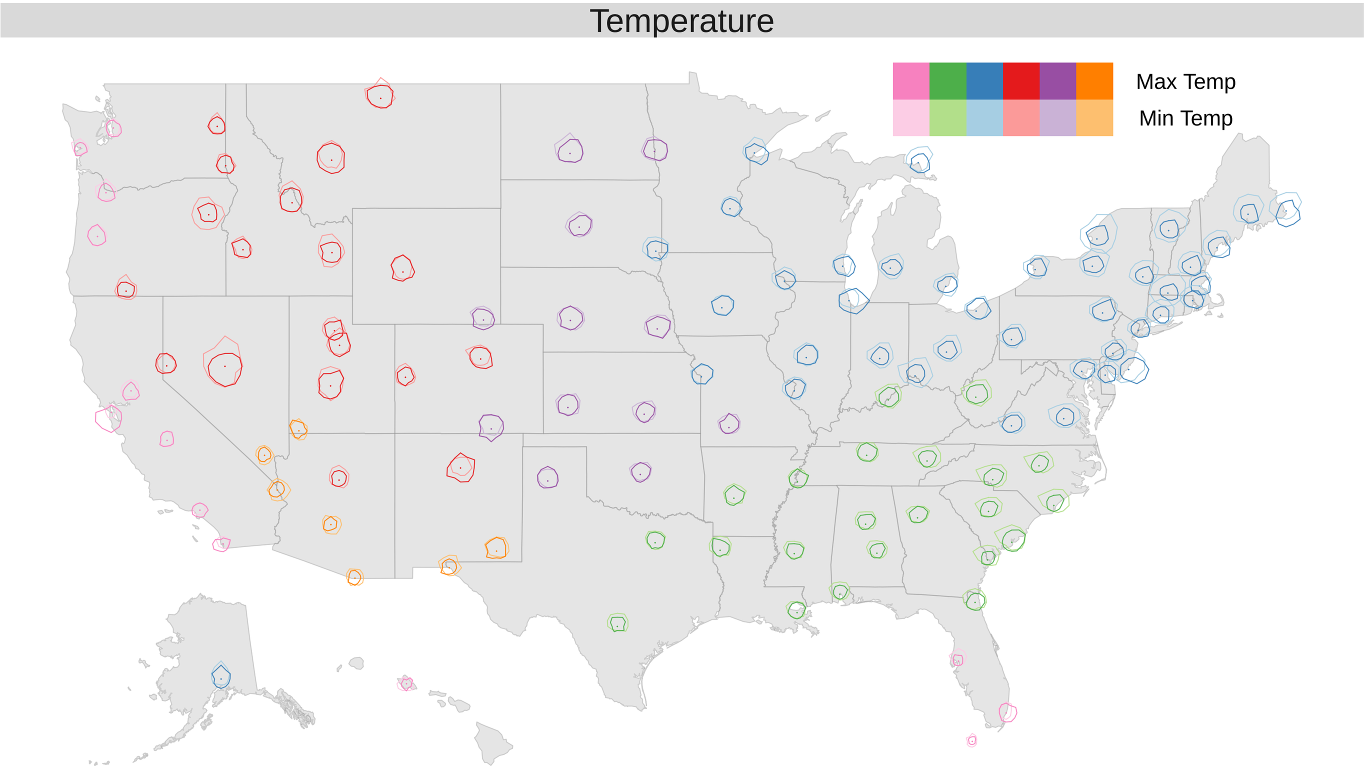
Northeast

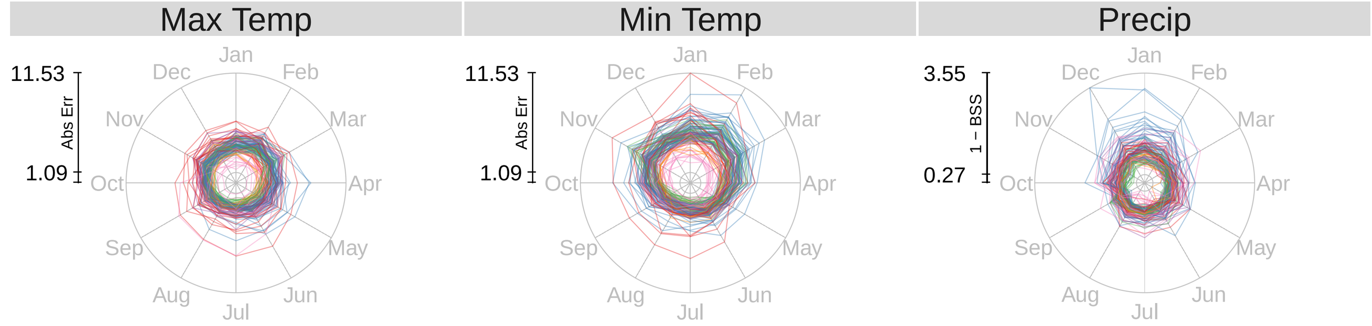
Southwest

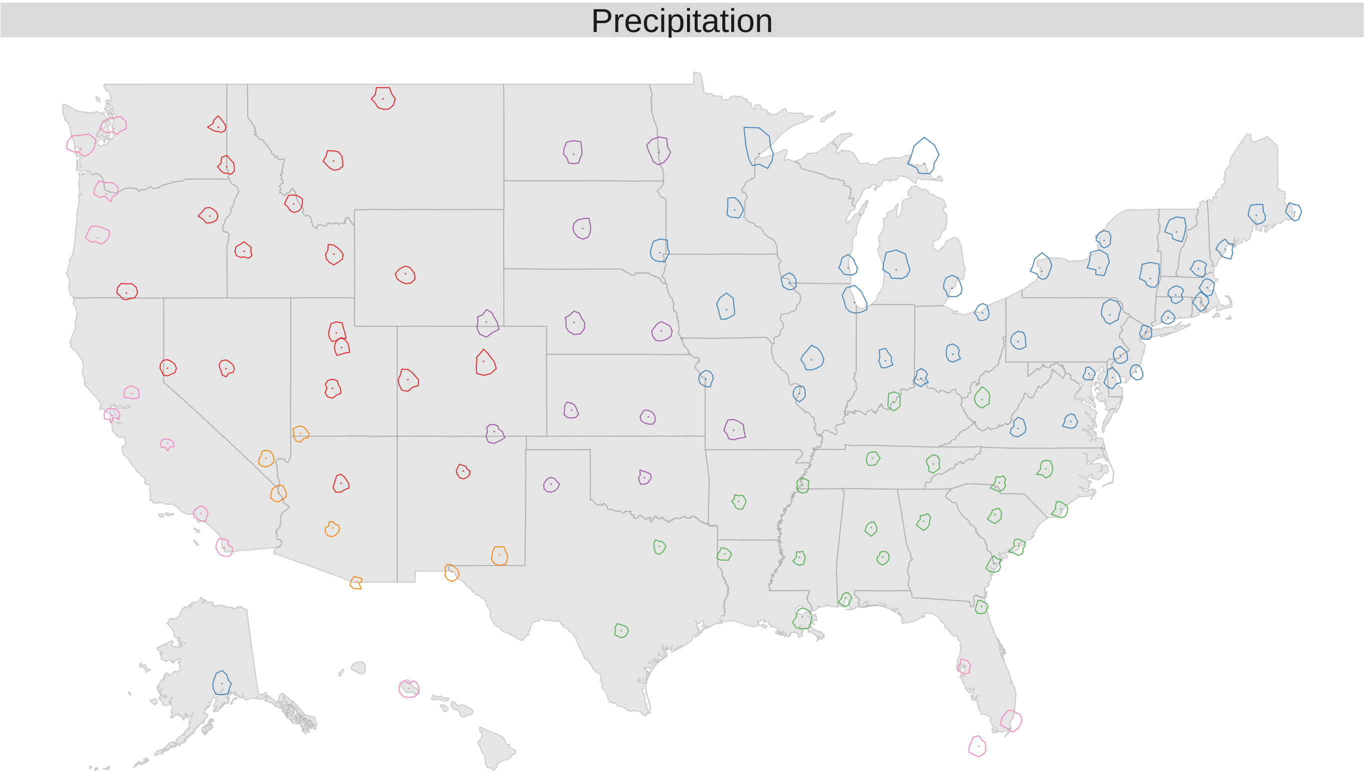




Seasonality







San Francisco, CA

Great Lakes

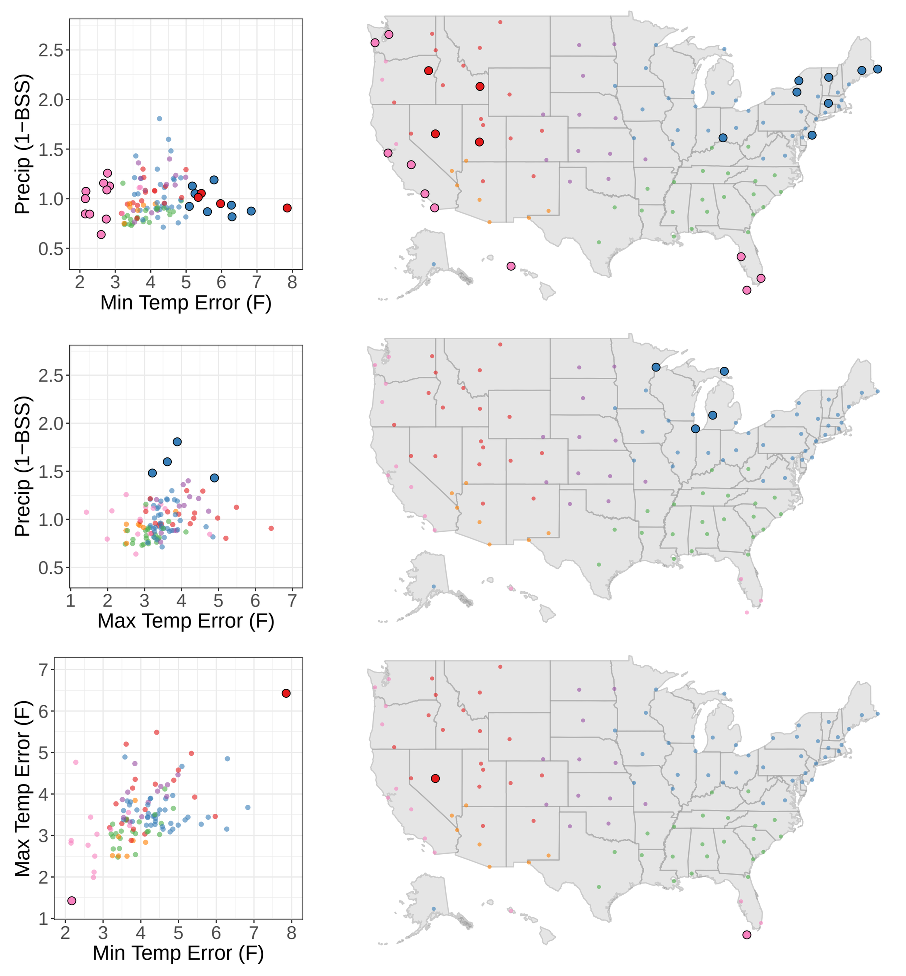
Trends & Outliers

Cali-Florida

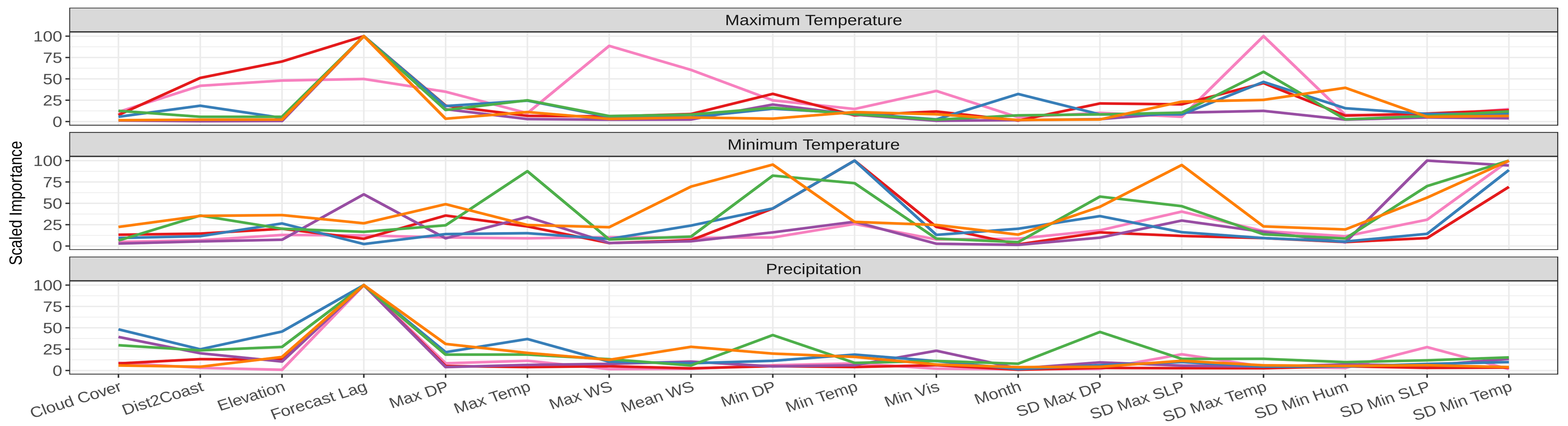
New England

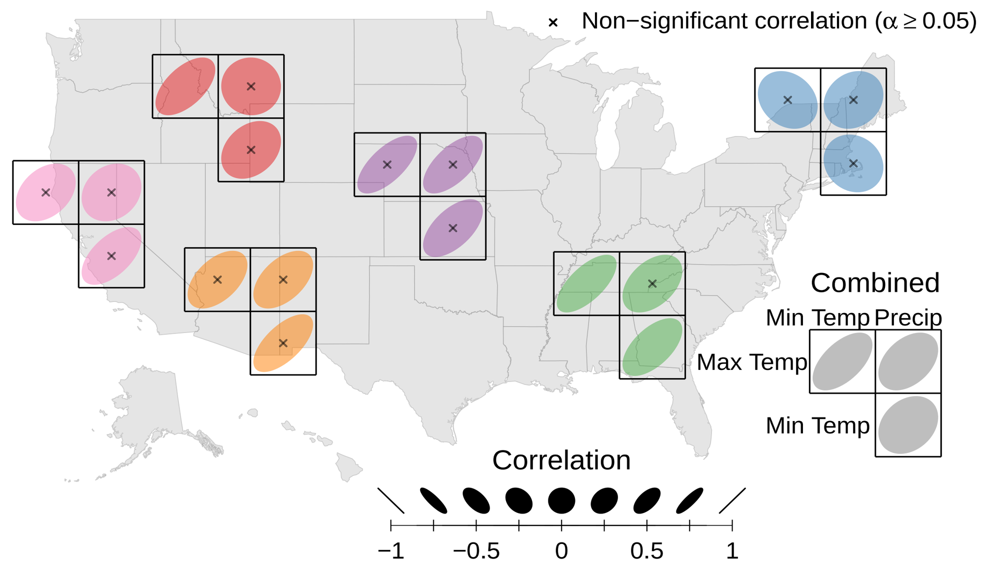
Key West, FL

Austin, NV



Importance & Correlations





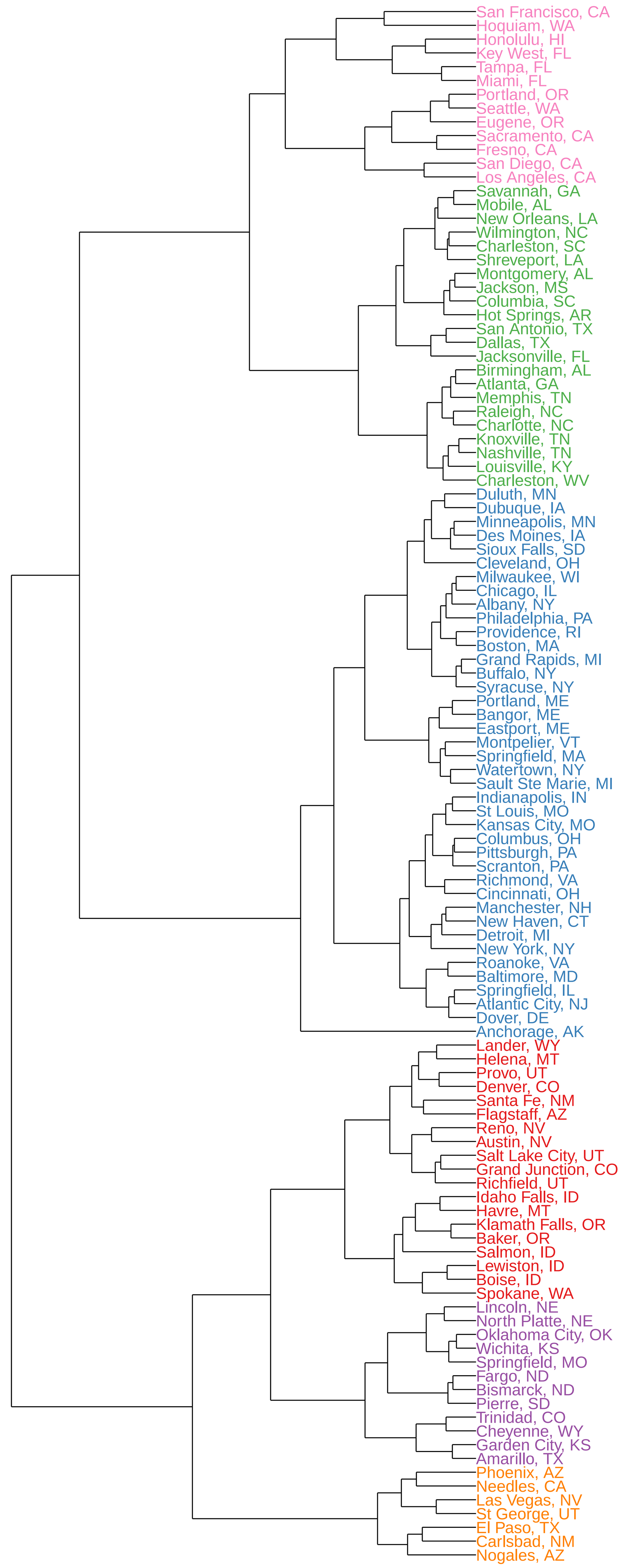
Conclusions

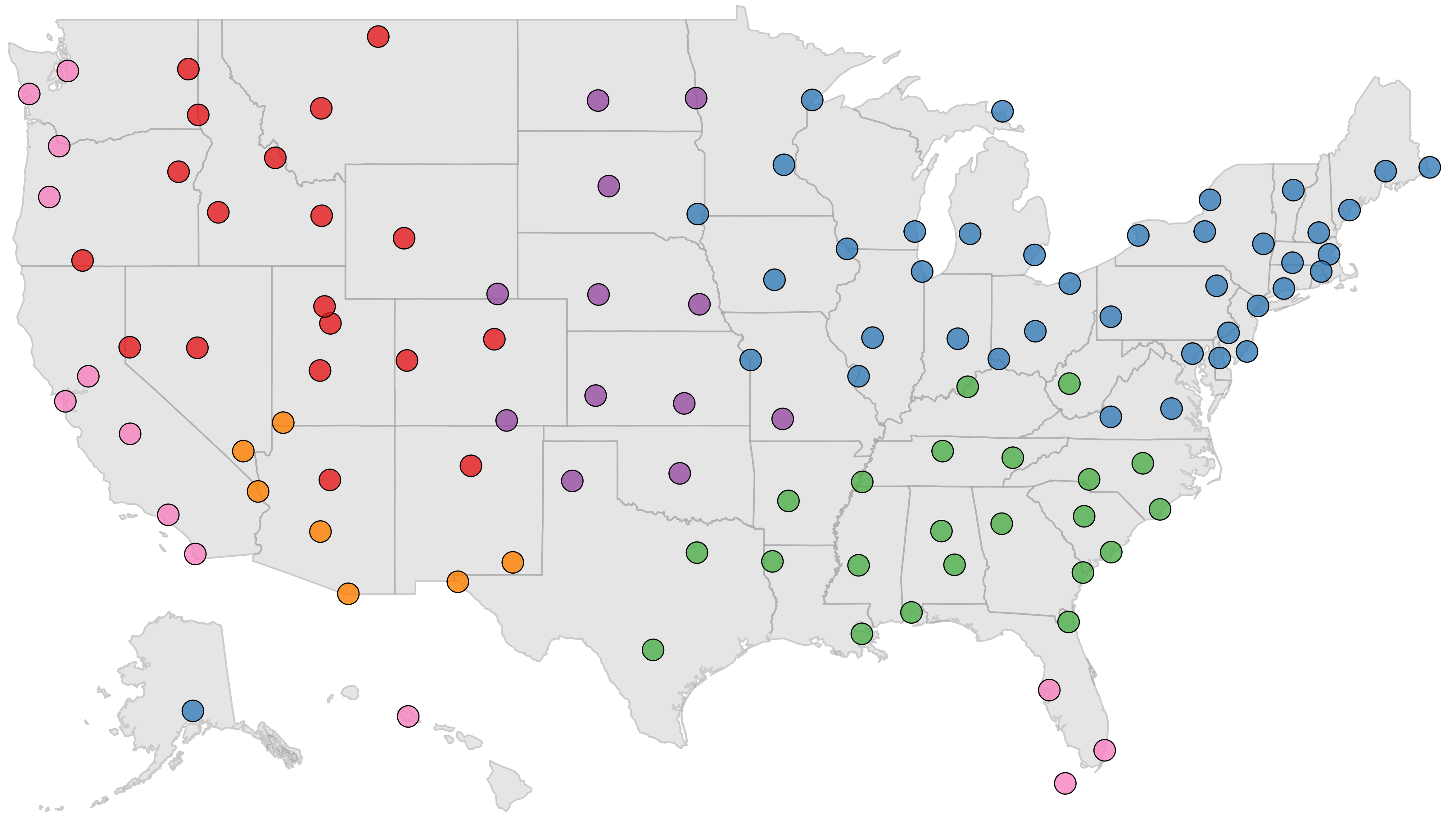
References

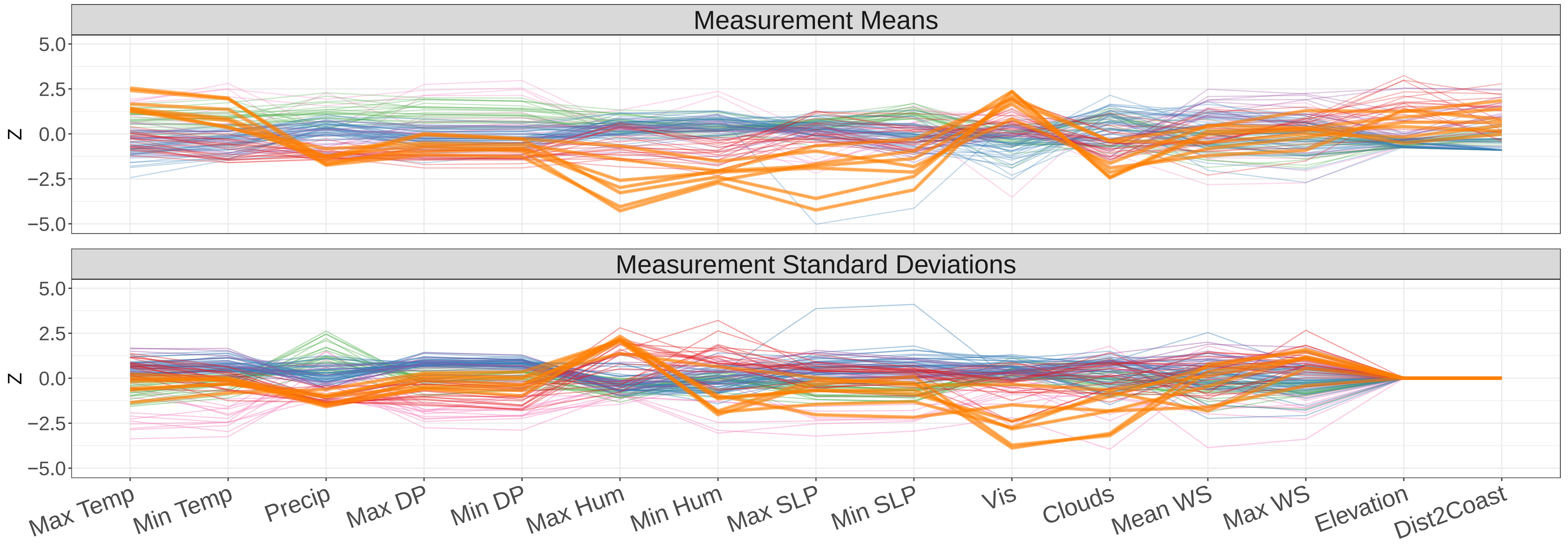
R Tools

Can we cluster the U.S. into regions based on weather measurements?

Overview | Clusters | Seasonality | Trends & Outliers | Importance & Correlations | Conclusions







Cluster Analysis

- Hierarchical clustering using **Ward's method** and **Euclidean distance**.
- Weather stations cleanly cluster into 6 weather regions:
  - Cali-Florida
  - Southeast
  - Northeast
  - Intermountain West
  - Midwest
  - Southwest
- Parallel coordinate plot of weather variables shows distinct weather patterns within each region (**See the app**).

How does forecast error change by cluster and by season?

Error Variables

- Absolute Maximum Temperature Error
- Absolute Minimum Temperature Error
- 1 - Brier Skill Score (BSS) for Precipitation

“The coldest winter I ever spent was a summer in San Francisco.”
- attr. Mark Twain

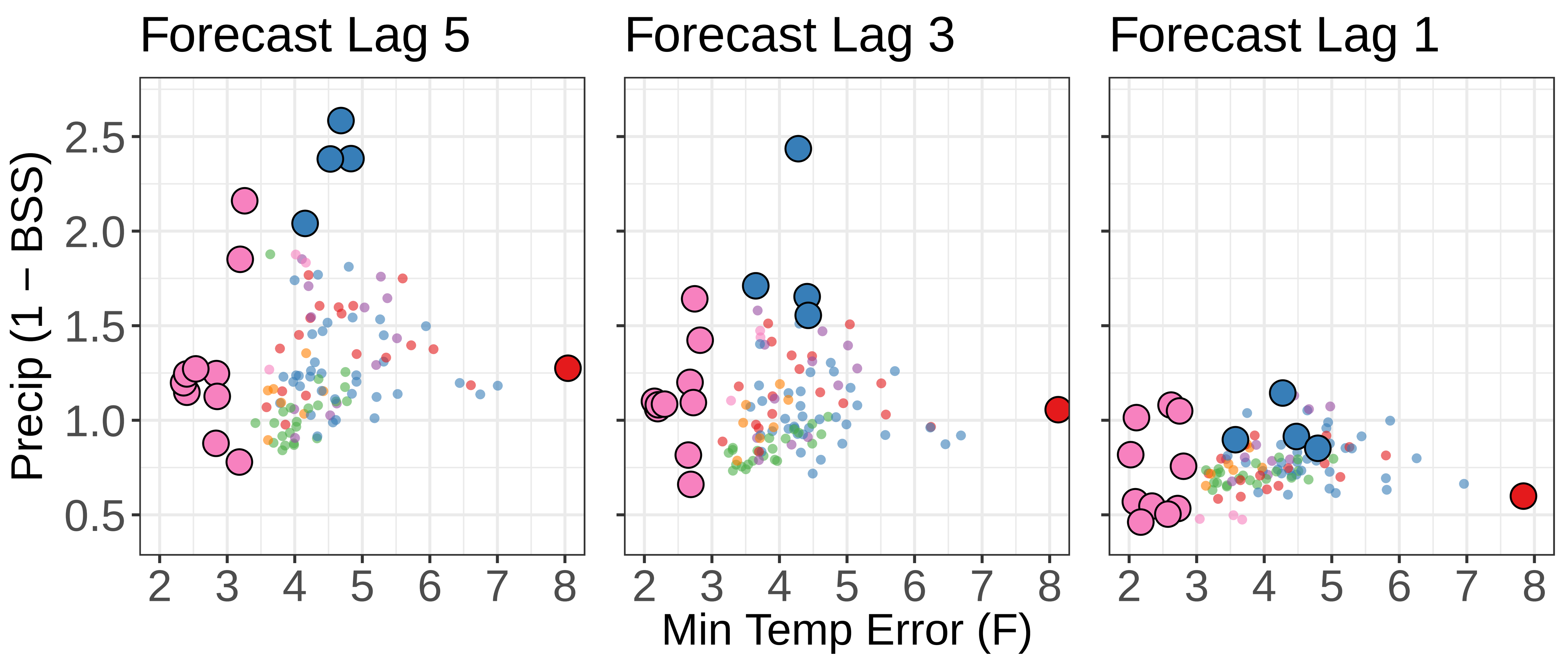
Global Trends

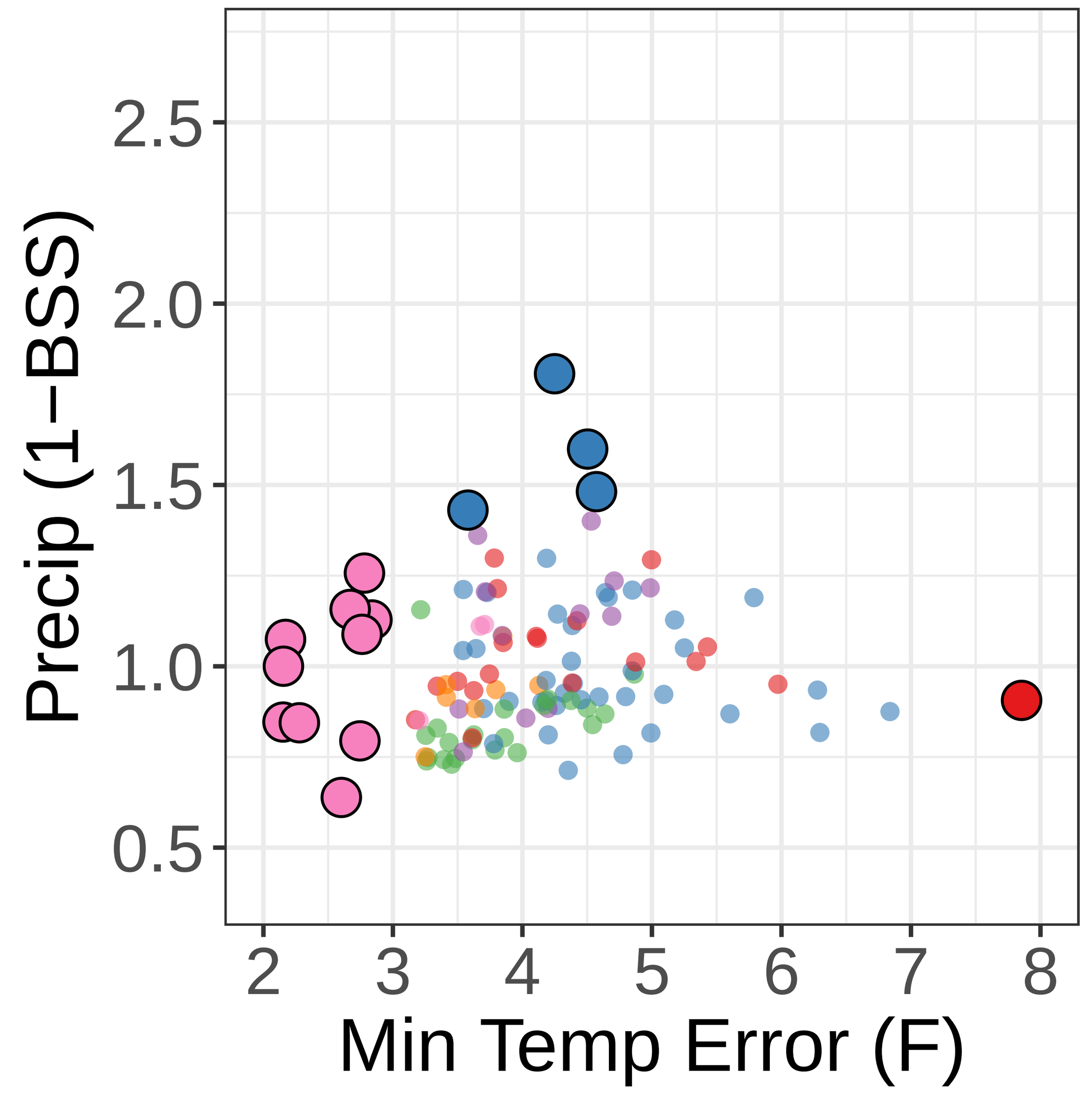
- Distinct seasonality in forecast accuracy.
- Max temp usually more accurate than min temp.
- Forecast accuracy improves north to south.

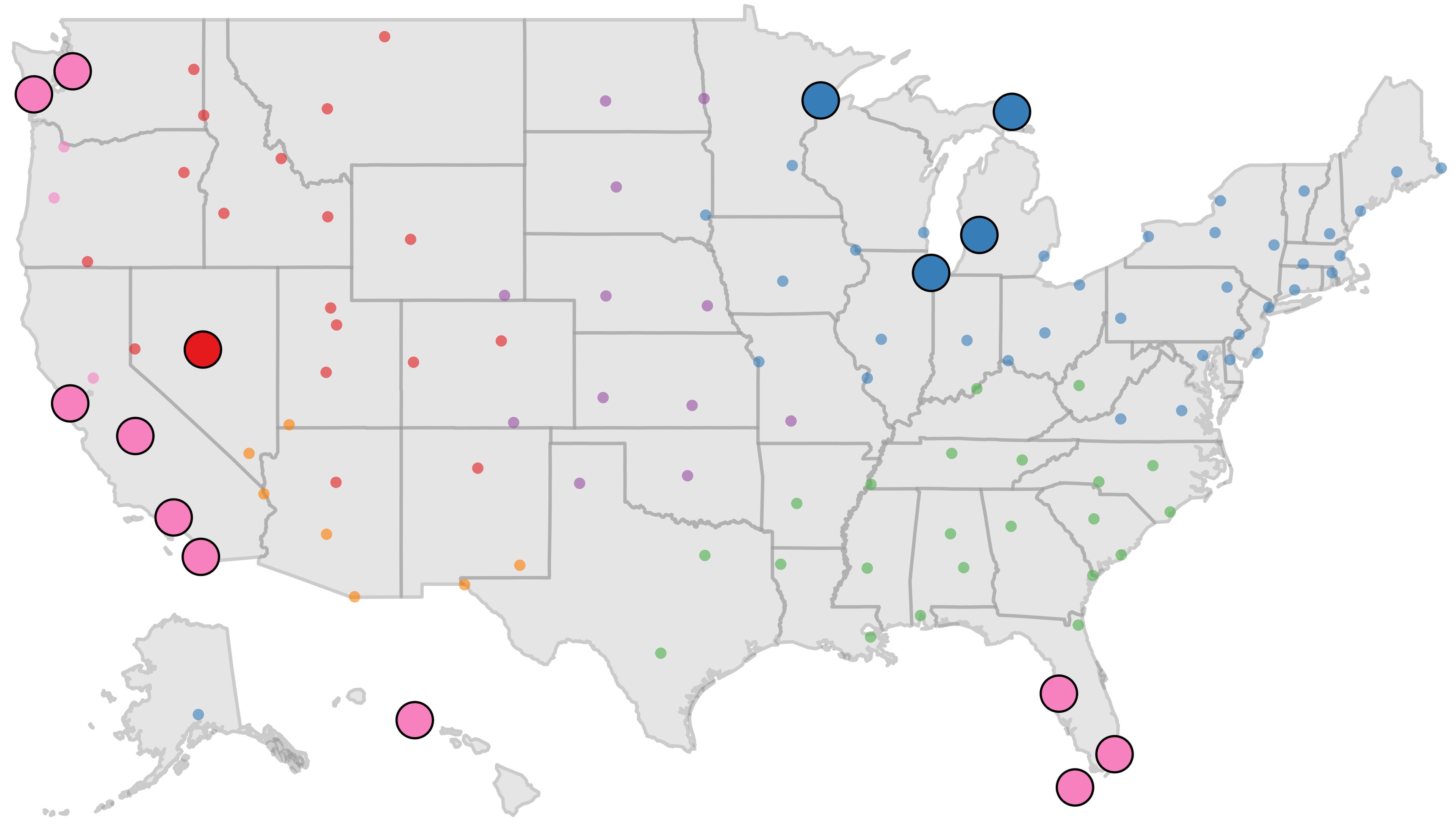
Local Trends

- The worst precip predictions are in the **Great Lakes** in winter.
- **San Francisco, CA**, predicts max temp well only in winter.
- **Austin, NV**, has the worst forecasts for both max and min temp.

Who are the winners and losers in terms of overall forecast accuracy?



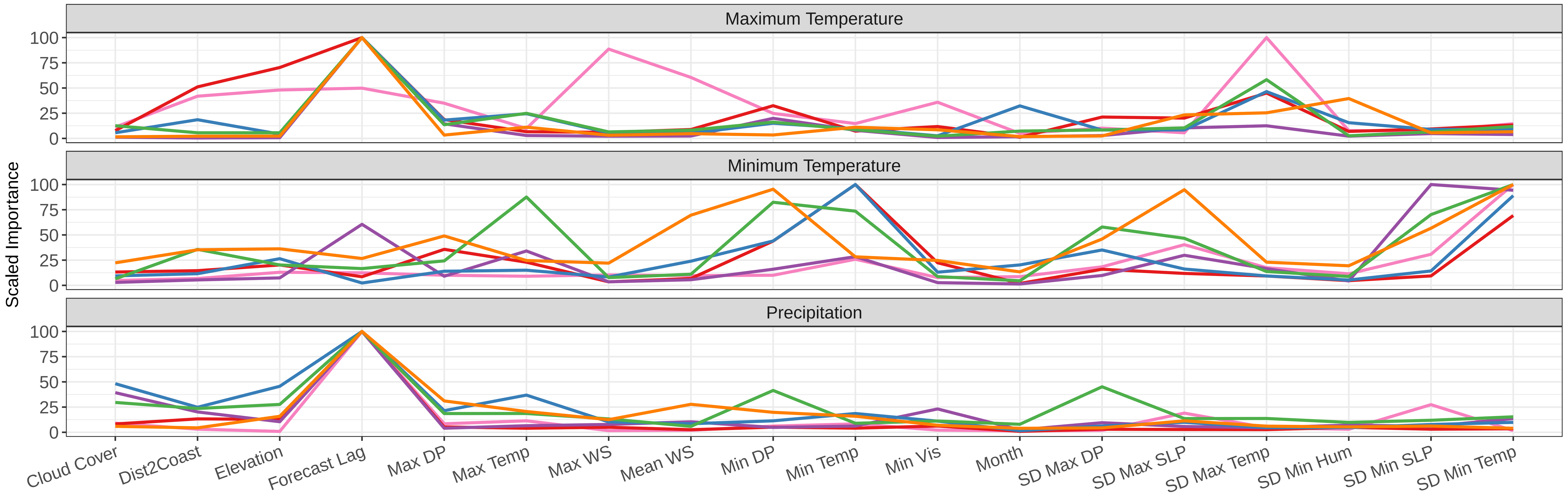


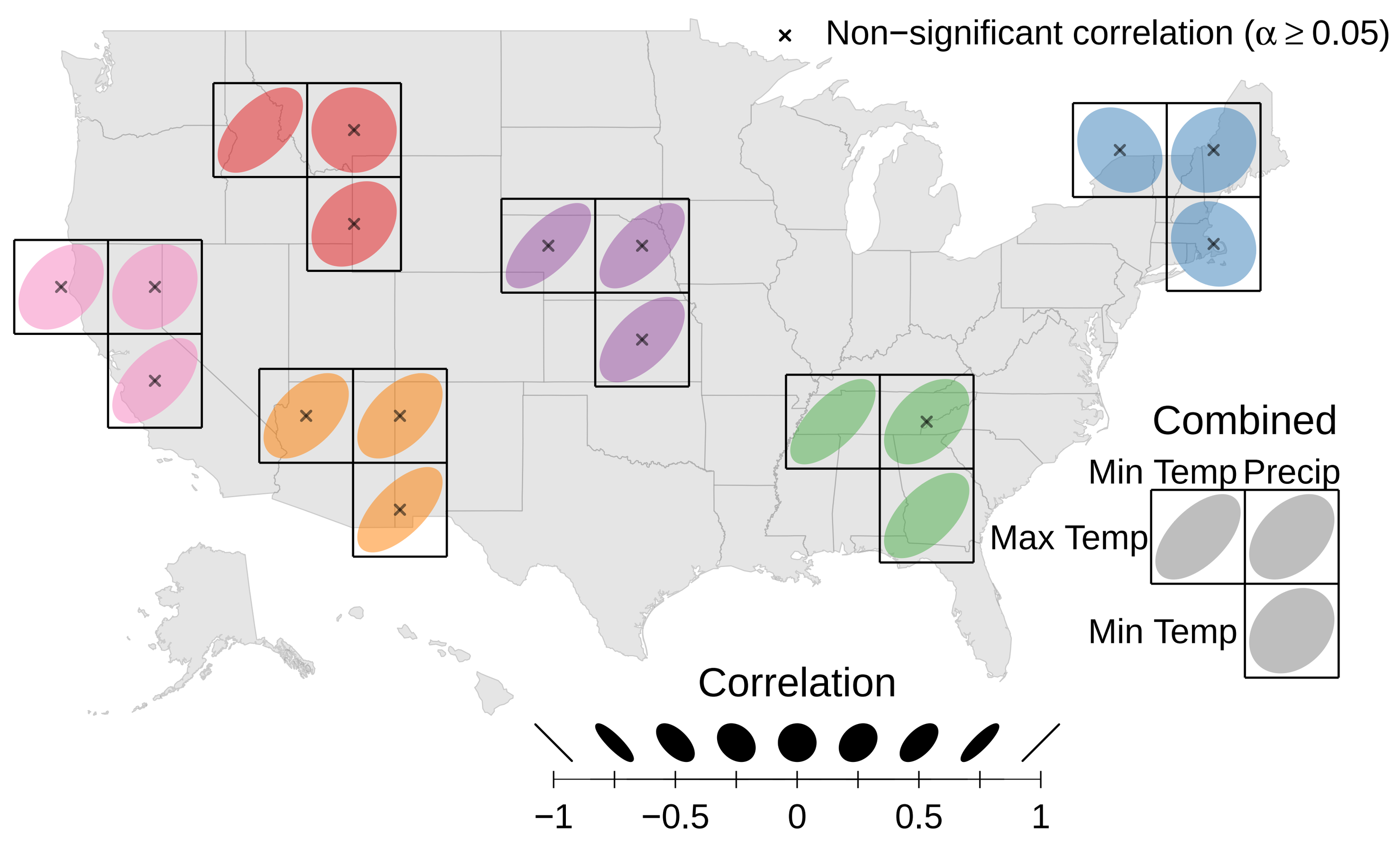


Trends & Outliers

- Previously seen poor forecasts for precip in the **Great Lakes** area and in min and max temp in **Austin, NV**, are apparent in these plots.
- Precip forecasts improve considerably as forecast lag decreases.
- **Cali-Florida** predicts min temp exceptionally well regardless of forecast lag.
- **See the app** for further interaction with forecast lag and trends.

Which variables are important in determining forecast error? How do error variables correlate?





Importance and Correlation

- **Random Forests** used to determine **importance** of each weather variable to forecast errors.
- Forecast lag most important in precip and important for max temp error with the exception of Cali-Florida, but not very important for min temp error.
- Important variables vary largely for each region with min temp error, but also differ with max temp error.
- **Correlations** between error variables differ for each region.

What did we learn?

Conclusions

- United States cleanly clusters into weather regions.
- Forecast error patterns differ by region, by season, and are related to regional climate characteristics.
- Forecasting anomalies exist within each region (**Great Lakes**; **San Francisco, CA**; **Austin, NV**).
- Forecast lag is the most important variable for precip across all regions, but important variables for max and min temp vary across regions.
- Correlations between error variables differ for each region.

Cali-Florida

Southeast

Northeast

Intermountain West

Midwest

Southwest

References

[1] A. Unwin, "Requirements for interactive graphics software for exploratory data analysis," *Computational Statistics*, vol. 14, no. 1, pp. 7–22, 1999.

[2] H. Wickham, H. Hofmann, C. Wickham, and D. Cook, "Glyph-maps for visually exploring temporal patterns in climate data and models," *Environmetrics*, vol. 23, no. 5, pp. 382–393, 2012.

[3] C. Nolte, "The story of the San Francisco summer is a bit foggy." https://www.sfchronicle.com, August 2016.

[4] R. W. Scott and F. A. Huff, "Impacts of the Great Lakes on regional climate conditions," *Journal of Great Lakes Research*, vol. 22, no. 4, pp. 845–863, 1996.

[5] N. Silver and R. Fischer-Baum, "Which city has the most unpredictable weather?." https://fivethirtyeight.com, December 2014.

[6] J. Cohen, K. Pfeiffer, and J. A. Francis, "Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States," *Nature Communications*, vol. 9, no. 1, p. 869, 2018.

[7] "Austin, Nevada: So much to do.." http://austinnevada.com.

[8] A. P. Weigel, M. A. Liniger, and C. Appenzeller, "The discrete Brier and ranked probability skill scores," *Monthly Weather Review*, vol. 135, no. 1, pp. 118–124, 2007.

[9] D. Murdoch and E. Chow, "A graphical display of large correlation matrices," *The American Statistician*, vol. 50, no. 2, pp. 178–180, 1996.

R Tools

- fields (D. Nychka et al. 2015)
- fiftystater (W. Murphy 2016)
- geosphere (R. Hijmans, 2016)
- ggforce (T. Pedersen 2018)
- gridExtra (A. Baptiste, 2017)
- latex2exp (S. Meschiari, 2015)
- reshape2 (H. Wickham, 2007)
- mapproj (D. McIlroy et al. 2017)
- randomForest (A. Liaw and M. Wiener, 2002)
- RColorBrewer (W. Neuwirth, 2014)
- reshape2 (H. Wickham, 2007)
- rgbif (S. Chamberlain, 2017)
- rgdal (R. Bivand et al. 2018)
- sp (E. Pebesma and R. Bivand, 2013)
- tidyverse (H. Wickham, 2017)
- weatherData (R. Narasimhan, 2017)