Data & Methodology

We used the following datasets:

- 1981–2010 U.S. Climate Normals from the National Centers for Environmental Information at the National Oceanic and Atmospheric Administration (NOAA)
  - Hourly normals: Cloudiness (percent clear, few clouds, scattered clouds, broken clouds, and overcast), dew point, heat index, pressure, temperature, wind chill, and windiness (wind speed and percent calm)
  - Year-to-date precipitation normals: rainfall and snowfall
  - Station metadata (station ID, state, elevation, latitude, and longitude)
- Day length in minutes for each weather station location, as generated with the Astral package in Python 3.6.0
- County boundaries shapefile from the United States Census Bureau
- Mapping of weather station location to county, as generated with the FCC's Census Block Conversions API

We implemented a variety of tools in obtaining, transforming, mapping, and processing data. Python 3.6.0 was used for data manipulations, as well as to generate data for mean daylight minutes for each weather station (using the Astral package). Counties outside the 48 contiguous states were deleted from shapefiles using QGIS 2.18.7. GeoDa 1.8.16.4 was used to map points and aerial units, as well as to select counties in which a weather station was located. Weather stations were mapped to counties using the FCC's Census Block Conversions API (via a script written in Python). County classification was done in RStudio 1.0.136 (running R 3.3.2).

We computed a total of 445 climate, daylight, and elevation variables. Descriptive statistics for aggregate measures are displayed in Table 1 below.

Climate data

Raw climate data were collected at weather stations located across the 48 contiguous states from 1981–2010. Data were collected hourly or daily, depending on the measure.

Hourly normals: Data available as hourly normals consist of 24 values per day for every day of the calendar year. Each value represents the average value for that hour for the years 1981–2010. For each station, hourly normals were averaged across each month; the minimum and maximum per-month values were also recorded for a selection of conditions (namely: each of the cloudiness measures, dew point, pressure, temperature, and the windiness conditions). The maximum heat index value was recorded, as was the minimum wind chill value.

Year-to-date normals: Data available as year-to-date normals consist of daily values for every day of each month of the calendar year. Each value represents the average value for that day for the years 1981–2010. For each station, daily normals were averaged across each month.

Station metadata: We filtered out stations outside the 48 contiguous United States and used only stations that had data for all the conditions of interest (N=321). We mapped stations to counties; when there was more than one station per county, we used the station listed first in the station metadata file. The resulting 288 weather stations are mapped in Figure 1 below. The counties in which weather stations are located are mapped in Figure 2 below.

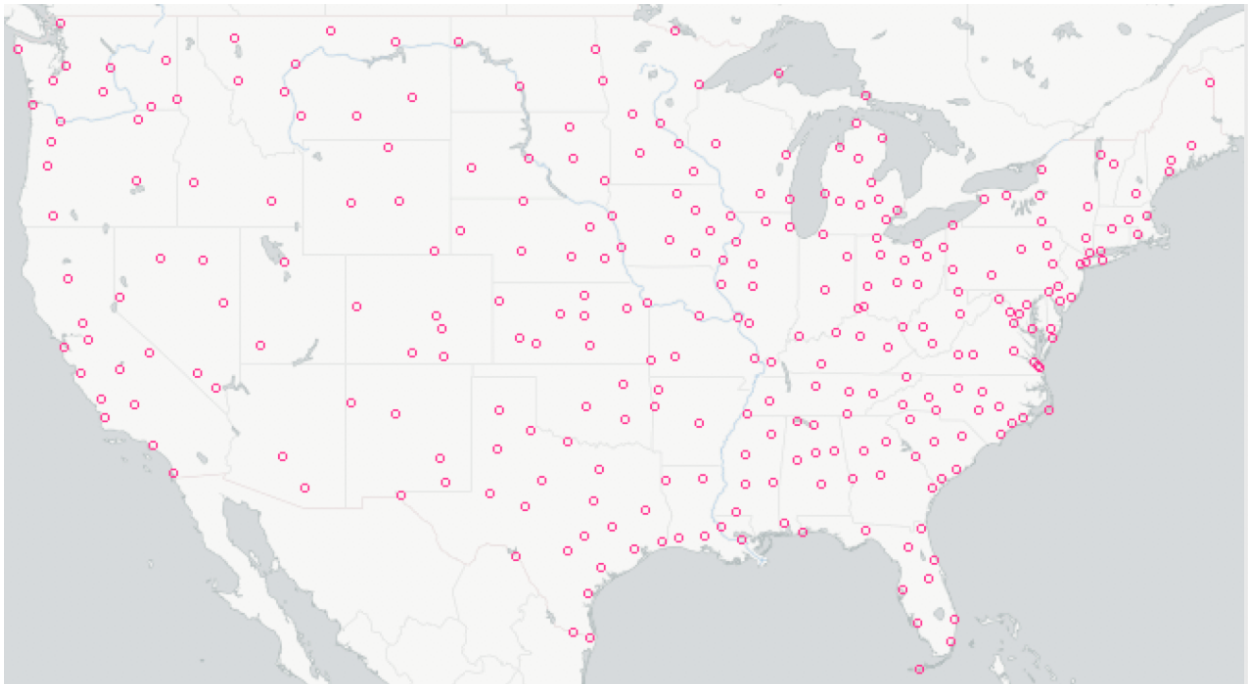

Figure 1: Weather stations (N=288).

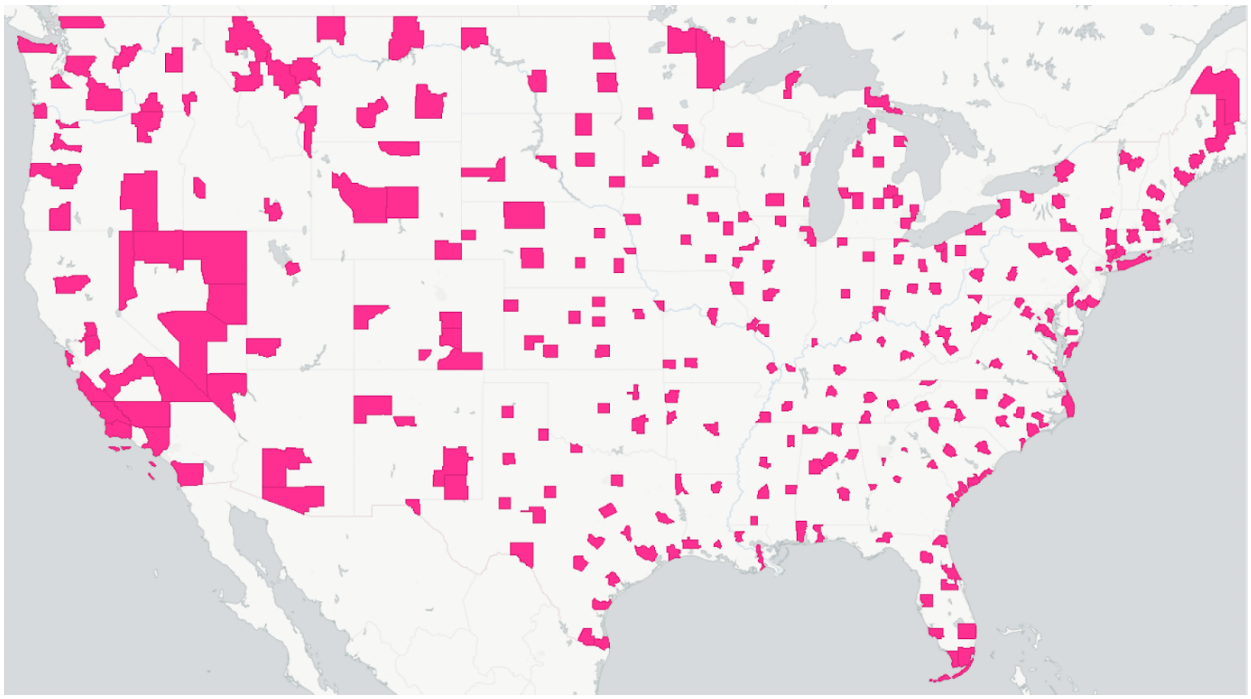

Figure 2: Counties in which weather stations (N=288) are located.

Elevation data

For each weather station, the location elevation in meters was recorded.

Daylight data

For each weather station location, daylight minutes (the length of the interval between sunrise and sunset) were calculated for each day between 1981 January 1 and 2010 December 31. Values were averaged to obtain a mean for each month of the calendar year.

	Min.	Max.	Median	Mean (s.d.)
Elevation: meters	1.200	2296.100	211.200	366.639 (456.977)
Monthly: Sunrise to sunset: Mean minutes	500.546	965.449	729.310	731.528 (113.617)
Monthly: Rainfall: Mean hundredths of inches	12.968	9343.097	1377.449	1740.148 (1399.732)
Monthly: Snowfall: Mean tenths of inches	0.000	2591.871	86.000	159.241 (204.236)
Monthly: Temperature: Min. degrees F	0.400	83.200	45.300	44.973 (16.208)
Monthly: Temperature: Max. degrees F	9.700	105.900	70.700	67.802 (17.467)
Monthly: Temperature: Mean degrees F	6.196	95.014	57.362	55.671 (17.182)
Monthly: Dew point: Min. degrees F	-3.900	74.700	38.200	38.600 (16.650)
Monthly: Dew point: Max. degrees F	2.800	76.400	48.400	47.710 (15.976)
Monthly: Dew point: Mean degrees F	-0.301	75.242	43.222	43.279 (16.406)
Monthly: Heat index: Max. degrees F	8.600	106.500	70.900	68.700 (18.465)
Monthly: Heat index: Mean degrees F	6.272	95.553	57.522	56.139 (17.676)
Monthly: Wind chill: Min. degrees F	-23.500	83.200	42.700	41.625 (19.307)
Monthly: Wind chill: Mean degrees F	-20.705	95.013	56.620	53.427 (19.951)
Monthly: Wind speed: Min. miles per hour	0.800	41.600	5.800	5.953 (2.638)
Monthly: Wind speed: Max. miles per hour	4.300	49.600	11.500	11.674 (2.822)
Monthly: Wind speed: Mean miles per hour	3.131	46.034	8.257	8.445 (2.611)
Monthly: Winds: Min. Percent Calm	0.000	38.800	1.100	2.077 (3.013)
Monthly: Winds: Max. Percent Calm	1.500	85.400	23.100	26.695 (15.894)
Monthly: Winds: Mean Percent Calm	0.644	47.558	10.301	12.640 (8.624)
Monthly: Barometric pressure: Min. millibars	1004.600	1022.000	1014.900	1014.745 (2.636)
Monthly: Barometric pressure: Max. millibars	1009.300	1029.100	1018.700	1018.770 (2.591)
Monthly: Barometric pressure: Mean millibars	1007.062	1024.775	1016.863	1016.780 (2.510)
Monthly: Clouds: Min. % Clear	0.000	88.500	18.100	19.503 (13.020)
Monthly: Clouds: Max. % Clear	9.300	98.700	47.300	47.950 (14.523)
Monthly: Clouds: Mean % Clear	3.235	95.430	32.580	32.927 (13.708)
Monthly: Clouds: Min. % Few	0.000	40.200	6.200	6.814 (3.549)
Monthly: Clouds: Max. % Few	5.000	56.300	19.500	20.733 (7.498)
Monthly: Clouds: Mean % Few	2.031	43.491	12.280	13.027 (5.058)
Monthly: Clouds: Min. % Scattered	0.000	18.900	2.300	2.831 (2.500)
Monthly: Clouds: Max. % Scattered	0.900	43.100	11.700	13.547 (7.512)
Monthly: Clouds: Mean % Scattered	0.600	30.303	6.114	7.179 (4.401)
Monthly: Clouds: Min. % Broken	0.000	29.300	7.600	8.206 (3.629)
Monthly: Clouds: Max. % Broken	3.700	63.300	25.200	26.485 (8.429)
Monthly: Clouds: Mean % Broken	1.019	45.864	15.541	16.435 (5.715)
Monthly: Clouds: Min. % Overcast	0.000	68.300	20.700	22.175 (13.057)
Monthly: Clouds: Max. % Overcast	2.200	85.200	39.600	40.030 (14.201)
Monthly: Clouds: Mean % Overcast	0.604	74.759	29.477	30.433 (13.552)

Table 1: Descriptive statistics for weather stations (N=288). "Monthly" indicates that there are twelve variables, one for each month of the calendar year, for a given measure.

Model

Based on the definition of climate as given by the Intergovernmental Panel on Climate Change (IPCC), we use monthly measures of a variety of climate variables averaged over 30 years (IPCC 2013). We also include elevation and mean daylength because both influence the way humans experience climate. The subjective experience of climate is as important as the objective reality of it for the purposes of classifying climate regions for use in social science research.

Clustering

Using GeoDa, neighbors were assigned based on Euclidean distance (derived from county centroids) and using a threshold[1] of 4.0. Neighbor counts ranged from 1 to 37 with a median at 16 and a mean at 16.9 (s.d. = 8.706). The connectivity graph can be seen in Figure 3 below.

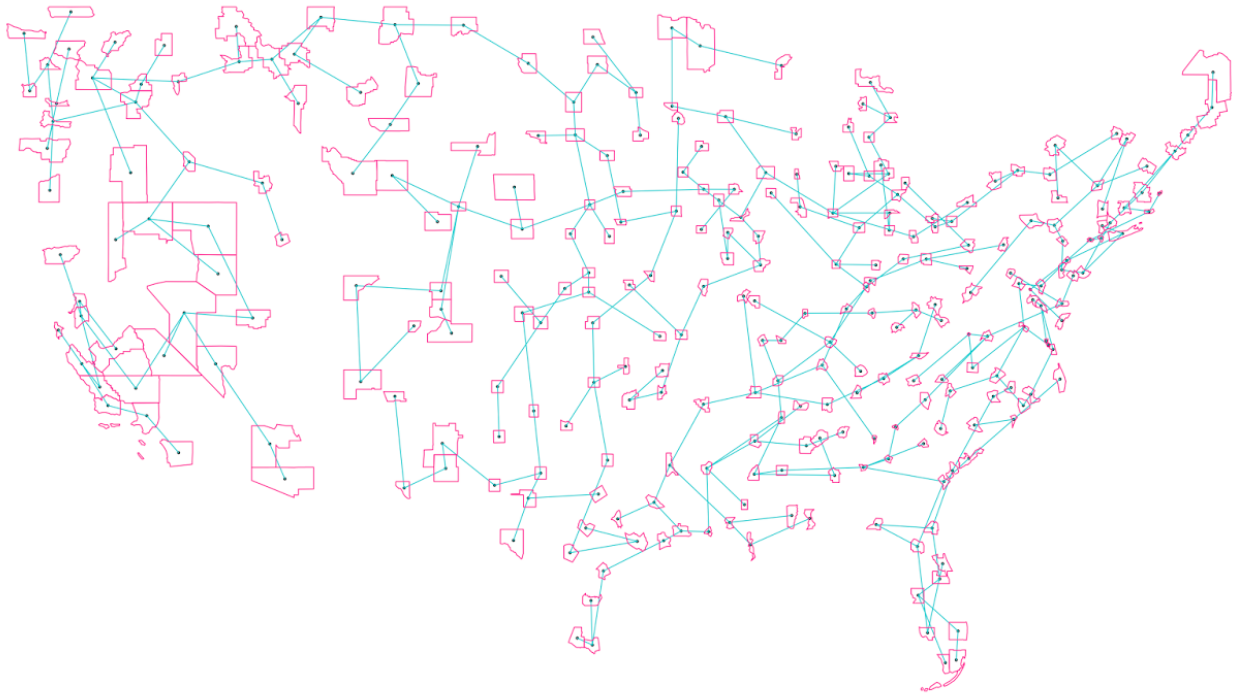

Figure 3: Connectivity graph. County borders are shown in pink; county centroids are black points; neighbor connections are blue lines.

The 445 climate, daylight, and elevation variables were scaled and centered. We computed clusters using the SKATER (Spatial Kluster Analysis by Tree Edge Removal) function from the spdep library in R and based on the scaled and centered measures.

Preliminary Results

Using SKATER, we compute 11 different models, classifying counties into categories with 5–15 cuts (6–16 categories, respectively) and visually assess the output of each model. By prioritizing both parsimony and accuracy, we find that the most balanced model is the one with 10 climate categories as displayed in Figure 4 below.

[1] The minimum threshold distance was 3.046018, but this produced two disjoint subgraphs; because the clustering algorithm requires a single graph, we increased the threshold to 4.0.

The 10-category model separates the Pacific Northwest (dark green) from the Pacific Southwest (pink) and inland West (yellow); in the 9-category model (displayed in Figure 5 below), however, the Pacific Northwest and the inland West are combined into one region (red). The 11-category model (shown in Figure 6 below) retains the Pacific Northwest as its own region and adds a small mid-Atlantic region (purple). The models with 8 or fewer categories (not shown) do not differentiate between regions known to have distinct climates; conversely, models with 12 or more regions (not shown) become increasingly more complex. Because our main goal is to develop a robust model of geographic climate regions, we find that the 10-category model to be the best because it is the simplest model that separates the Pacific Northwest, which is known to have a distinct climate, from the inland West.

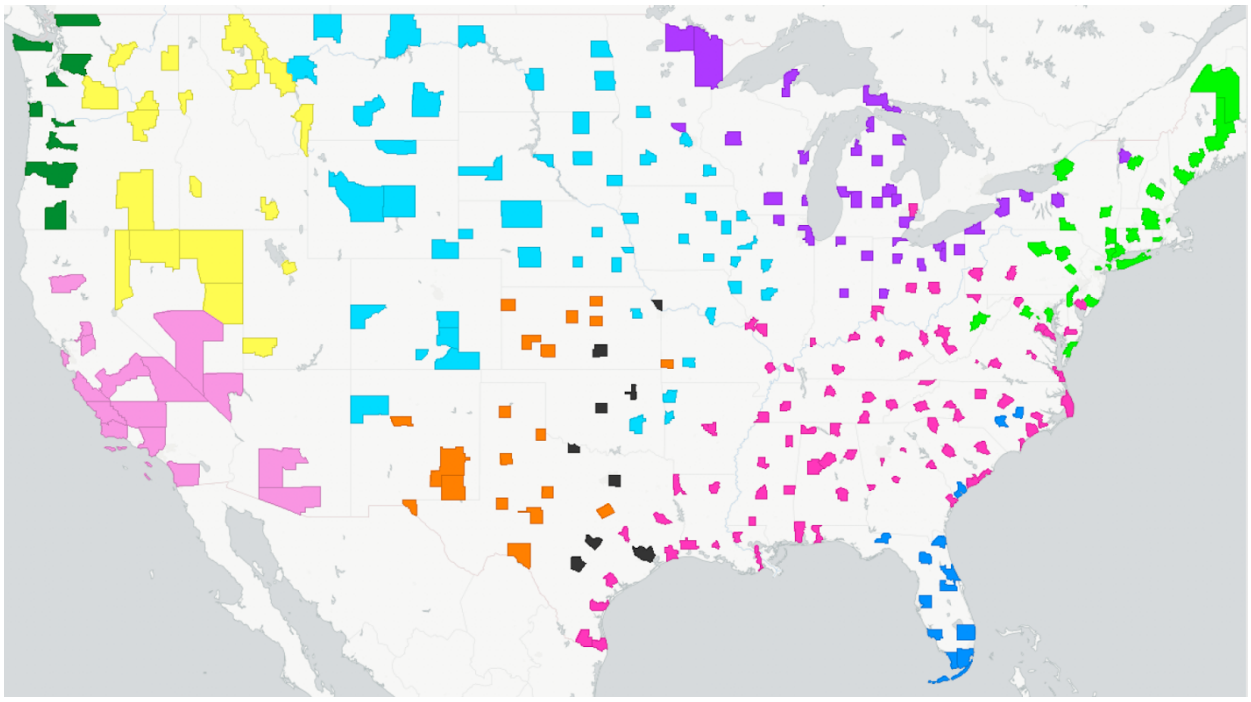

Figure 4: Counties classified into 10 climate categories (identified by color) with SKATER.

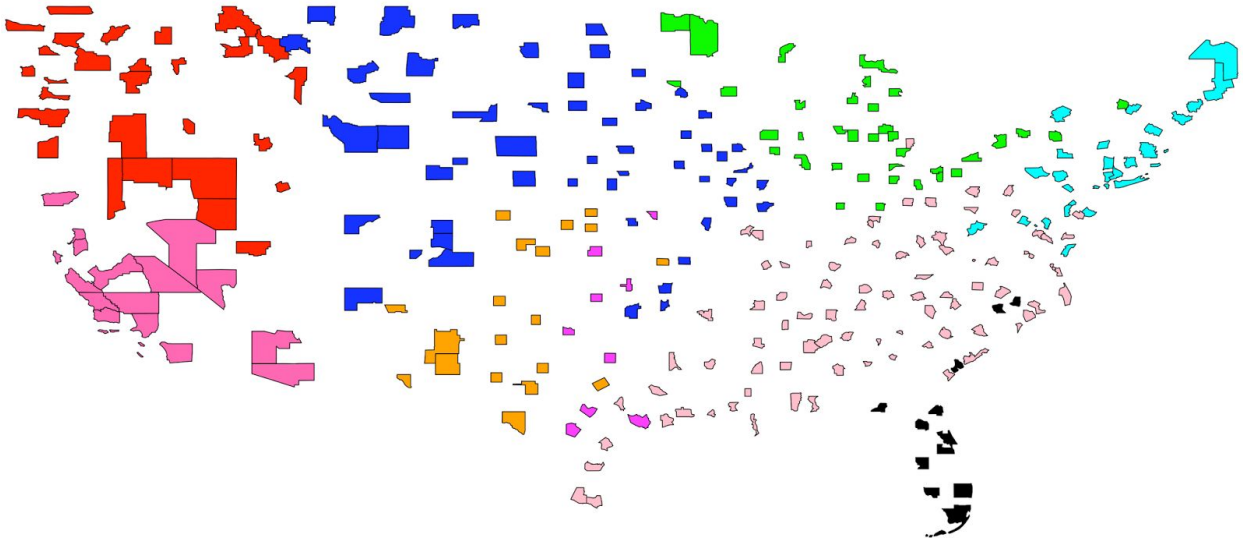

Figure 5: Counties classified into 9 climate categories (identified by color) with SKATER.

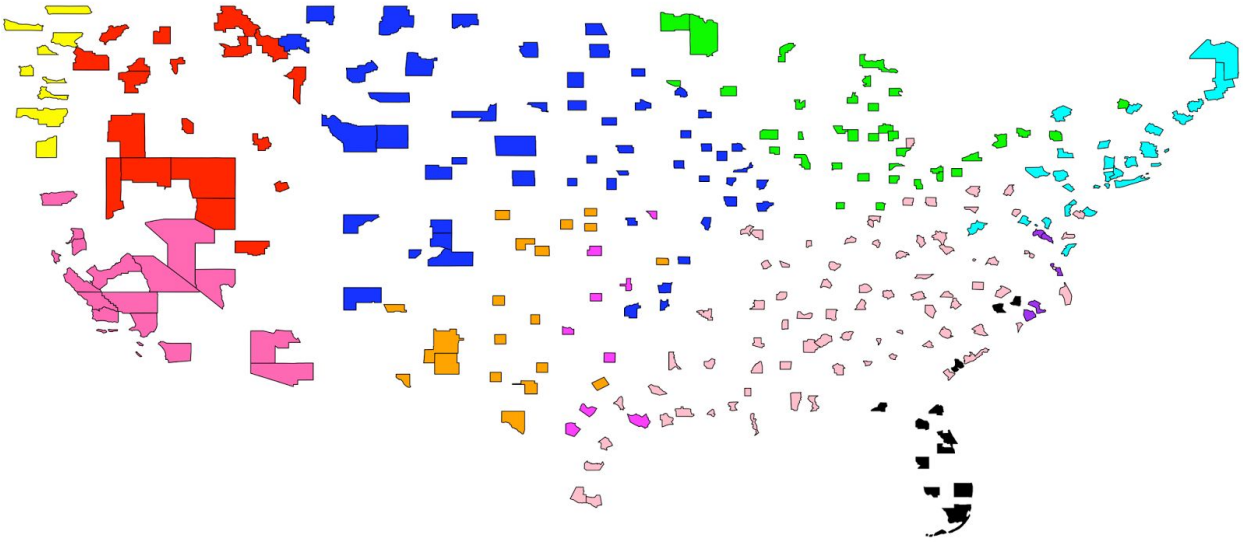

Figure 6: Counties classified into 11 climate categories (identified by color) with SKATER.

References

Federal Communications Commission. N.d. "Census Block Conversions API." fcc.gov/general/census-block-conversions-api

Intergovernmental Panel on Climate Change (IPCC). 2013. "Annex III: Glossary." *Climate Change 2013: The Physical Science Basis.* ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_AnnexIII_FINAL.pdf

Kennedy, Simon. 2017. "Astral v1.4." pythonhosted.org/astral

National Oceanic and Atmospheric Administration (NOAA). 2017. "1981–2010 U.S. Climate Normals." doi.org/10.7289/V5PN93JP

United States Census Bureau. N.d. "Cartographic Boundary Shapefiles - Counties: 2016." census.gov/geo/maps-data/data/cbf/cbf_counties.html