CS 6220 Data Mining Techniques — Final Project

Part: Exploratory Data Analysis (EDA): Including correlation heatmaps, scatter plots, and geospatial visualizations.

Introduction and Dataset Selection

For the **Final Project**, we selected a dataset from the **NOAA Daily Global Historical Climatology Network**, which contains daily U.S. weather measurements for the year **2017**. The dataset has been pre-processed to remove weather stations with sparse data. Each column represents different meteorological variables, with descriptions provided in the accompanying data file.

Through this exploration, I aim to answer compelling questions about **temperature trends, extreme weather events, and relationships between different meteorological variables**—all while deepening our understanding of the climate in places that have shaped my experiences.

Analysis Questions

1. How does temperature, precipitation, wind speed, snowfall vary across all states over time in 2017? Are there any seasonal patterns or unexpected anomalies?

We pre-processed the dataset using Pandas to ensure data consistency and quality before conducting our analysis. The preprocessing involved handling missing values, converting date formats, and filtering out incomplete or duplicate records.

- Removed missing values in critical fields such as date, geographic information (state, latitude, longitude, elevation), and key weather metrics (TMIN, TMAX, TAVG, AWND, WSF5, PRCP).
- Converted dates from YYYYMMDD format to YYYY-MM-DD to facilitate time-based analysis.
- Eliminated duplicate records based on station and date to ensure data integrity.
- Filtered out stations with sparse measurements, retaining only those with sufficient data coverage for meaningful insights.

The pre-processed data located in https://github.com/QingyuanWan/CS6220WeatherAnalysis/blob/main/data_pre_processing.ipynb

```
weather = ▸Array(71978) [Object, Object, Object, Object, Object, Object, Object, Object, Object, Object, Object, Object, Ol
```

station	state	latitude	longitude	elevation	date	TMIN	TMAX	TAVG	AWND	WDF5	WSF5
"GUAM INTL AP"	"GU"	"13.4836"	"144.7961"	"77.4"	"2017-03-12"	"71.06"	"87.08"	"80.06"	"4.47388"	"360.0"	"21.02723€
"KALISPELL GLACIER AP"	"MT"	"48.3042"	"-114.2636"	"901.3"	"2017-02-07"	"-0.76"	"22.1"	"13.64"	"3.802798"	"360.0"	"14.092722
"KALISPELL GLACIER AP"	"MT"	"48.3042"	"-114.2636"	"901.3"	"2017-03-30"	"37.04"	"53.96"	"44.24"	"4.026492"	"360.0"	"19.90876€
"KALISPELL GLACIER AP"	"MT"	"48.3042"	"-114.2636"	"901.3"	"2017-06-22"	"35.96"	"73.04"	"59.72"	"3.579104"	"360.0"	"19.01399"
"KALISPELL GLACIER AP"	"MT"	"48.3042"	"-114.2636"	"901.3"	"2017-07-25"	"53.06"	"87.08"	"71.6"	"6.039738"	"360.0"	"21.922012
"DENVER INTL AP"	"CO"	"39.8328"	"-104.6575"	"1650.2"	"2017-01-05"	"-1.84"	"5.18"	"3.56"	"4.921268"	"360.0"	"10.96100€

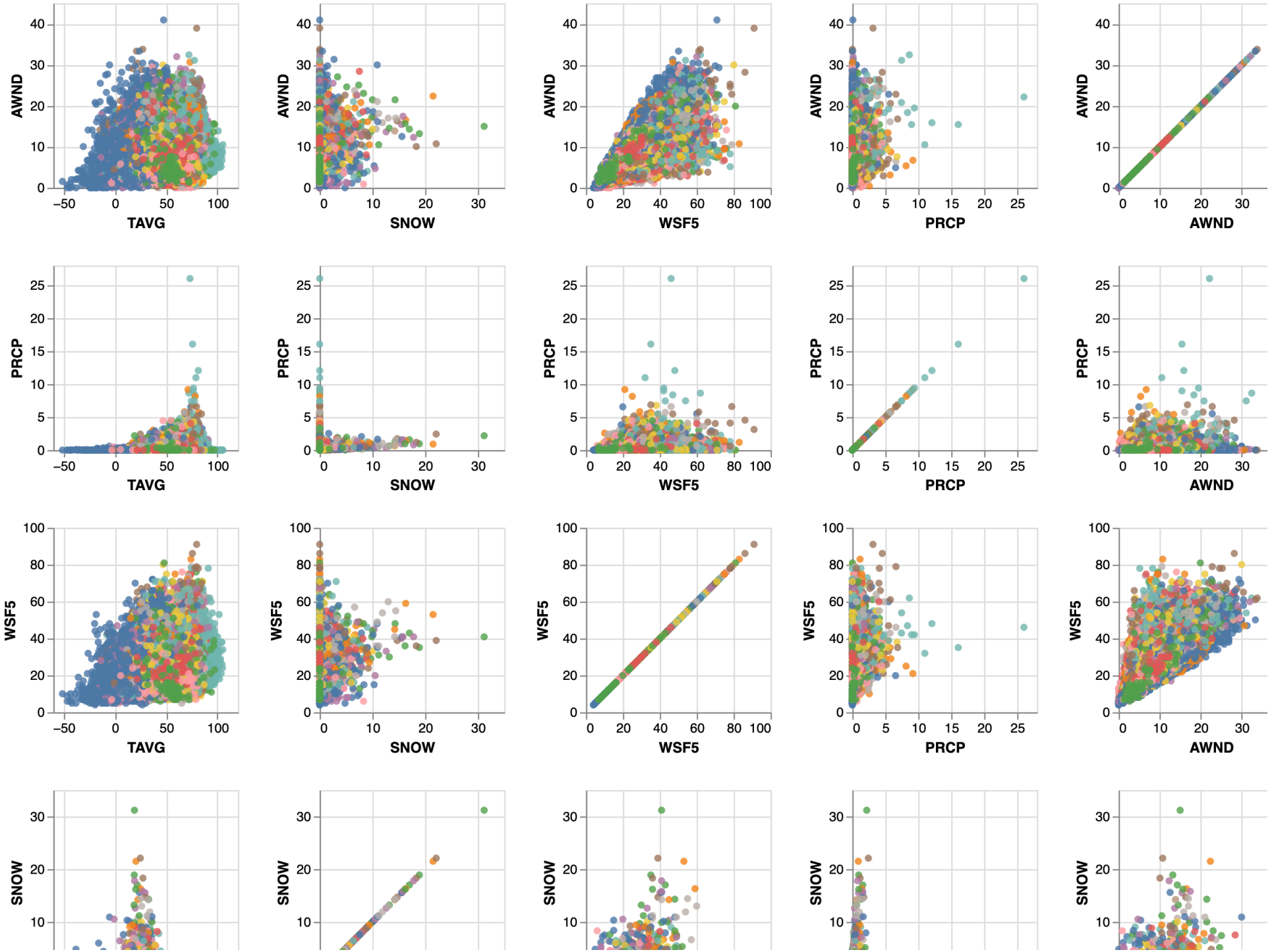
station	state	latitude	longitude	elevation	date	TMIN	TMAX	TAVG	AWND	WDF5	WSF5
"DENVER INTL AP"	"CO"	"39.8328"	"-104.6575"	"1650.2"	"2017-03-29"	"33.98"	"57.02"	"41.36"	"12.750558"	"360.0"	"29.974996
"DENVER INTL AP"	"CO"	"39.8328"	"-104.6575"	"1650.2"	"2017-04-29"	"29.12"	"31.1"	"30.2"	"16.105968"	"360.0"	"29.974996
"DENVER INTL AP"	"CO"	"39.8328"	"-104.6575"	"1650.2"	"2017-05-23"	"41.0"	"62.96"	"48.74"	"11.1847"	"360.0"	"34.001488
"DENVER INTL AP"	"CO"	"39.8328"	"-104.6575"	"1650.2"	"2017-07-03"	"53.06"	"91.94"	"73.22"	"9.618842"	"360.0"	"40.936002

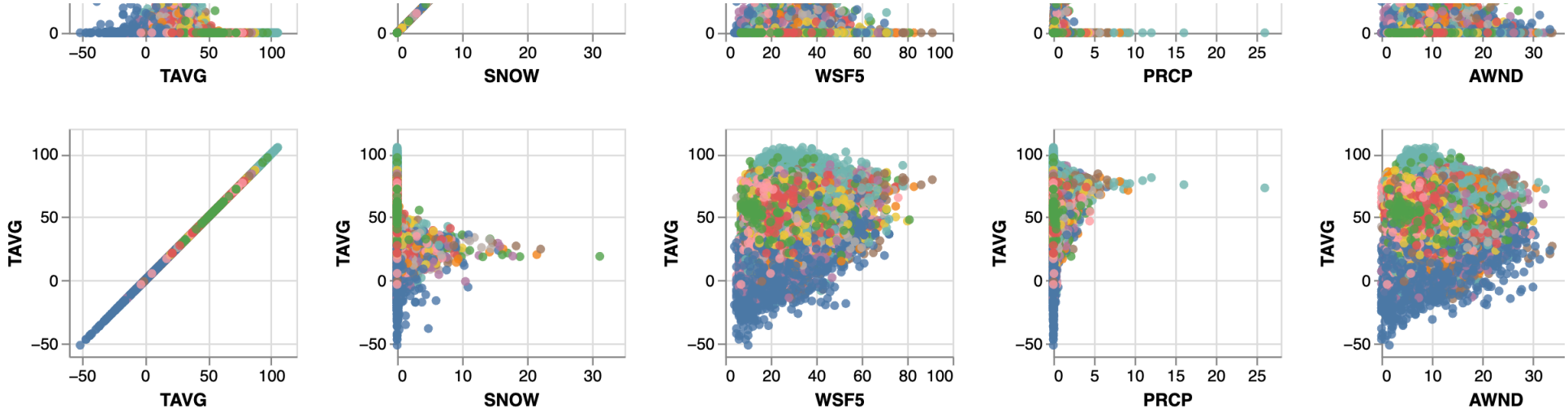
```
weatherForPlotMonth = ▸Array(71978) [Object, Object, Object, Object, Object, Object, Object, Object, Object, Object, Objec
```

Before answering my core research questions, I begin with an exploratory analysis of the dataset. The goal of this step is to roughly identify patterns, trends, and anomalies in the weather data for states over the course of 2017.

We visualized all relevant weather variable to capture the relationships between different climate factors, and plotted a scatter matrix where each variable (temperature, precipitation, wind speed, snowfall, etc.) is compared against others. This could helps identify potential correlations, outliers, and state-specific trends.

Climate Correlations Across all States in the US: Temperature, Precipitation, Wind, and Snowfall Patterns





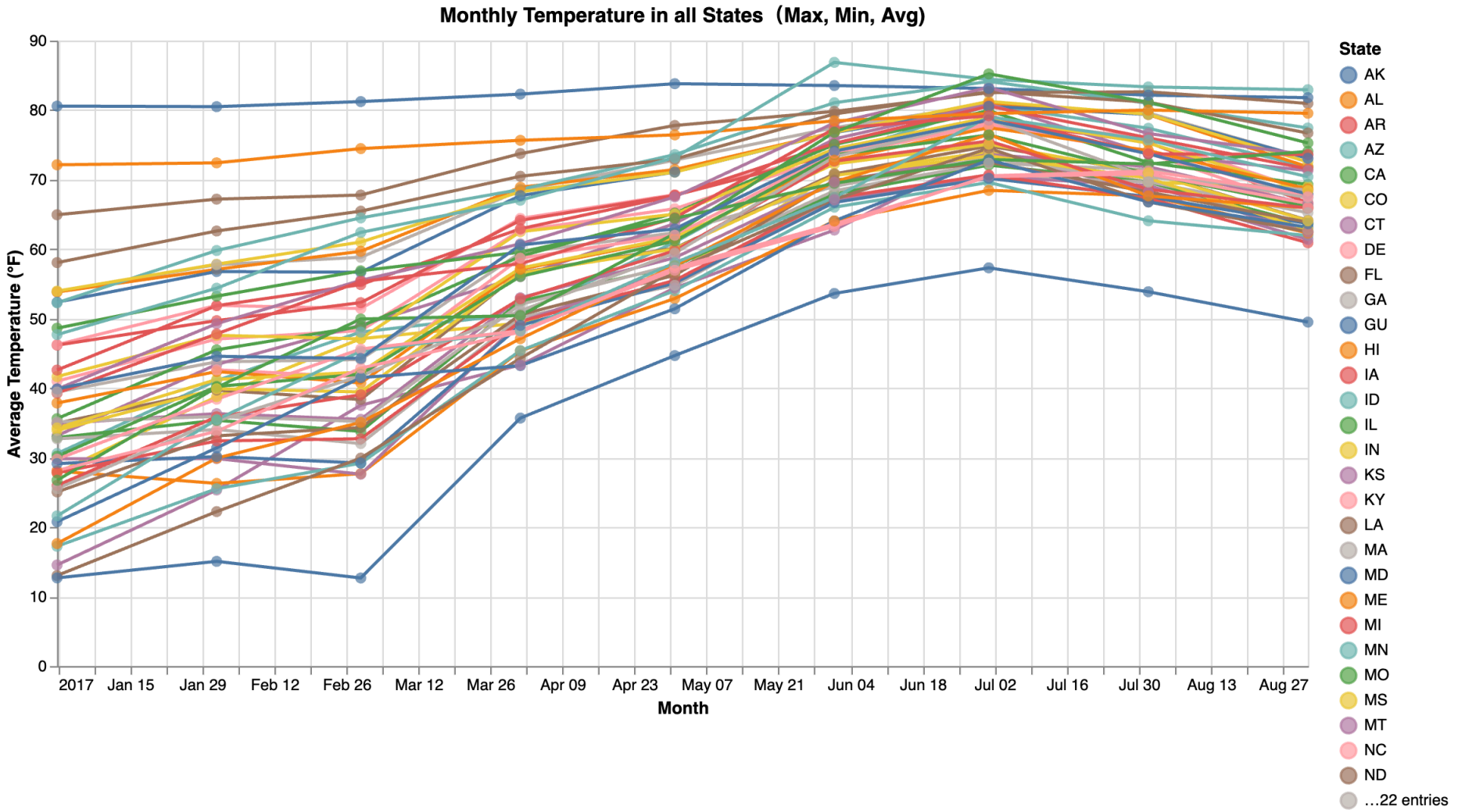
Time Series (Line Charts by Month)

I aggregated data by month and state to track trends over 2017:

Monthly Temperature (mean, max, min)

Findings:

- All states warm up from winter to summer, peaking around July/August.
- AK sees significant temperature drops in winter, down to 15°F or below.
- The average temperature in GU remains high year-round.

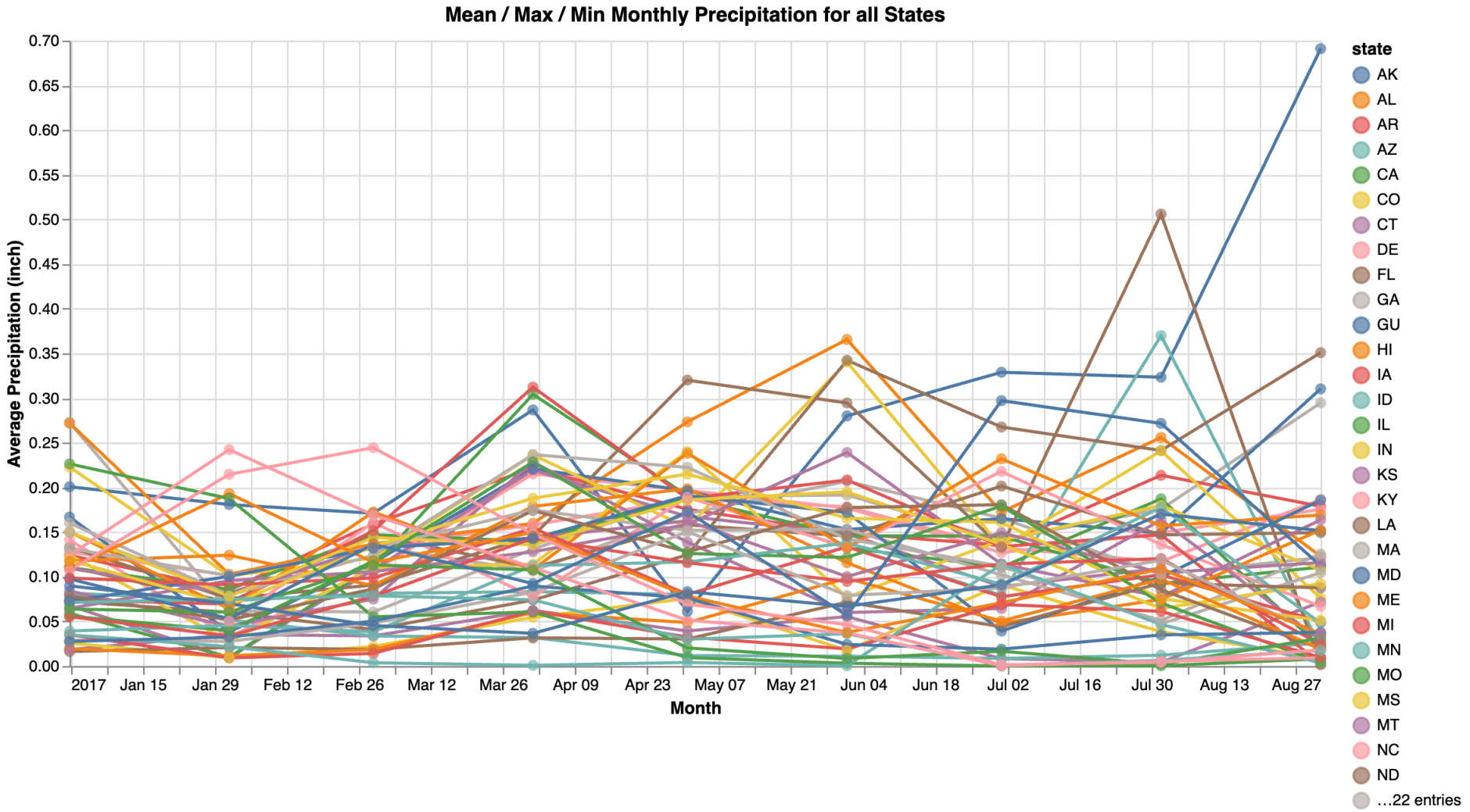


tempType TAVG

Monthly Precipitation

Findings:

- We can see TX got maximum 26 inches precipitation cause TX expericnce storm Cindy(https://www.weather.gov/mob/cindy)
- (Click `max_PRCP`)

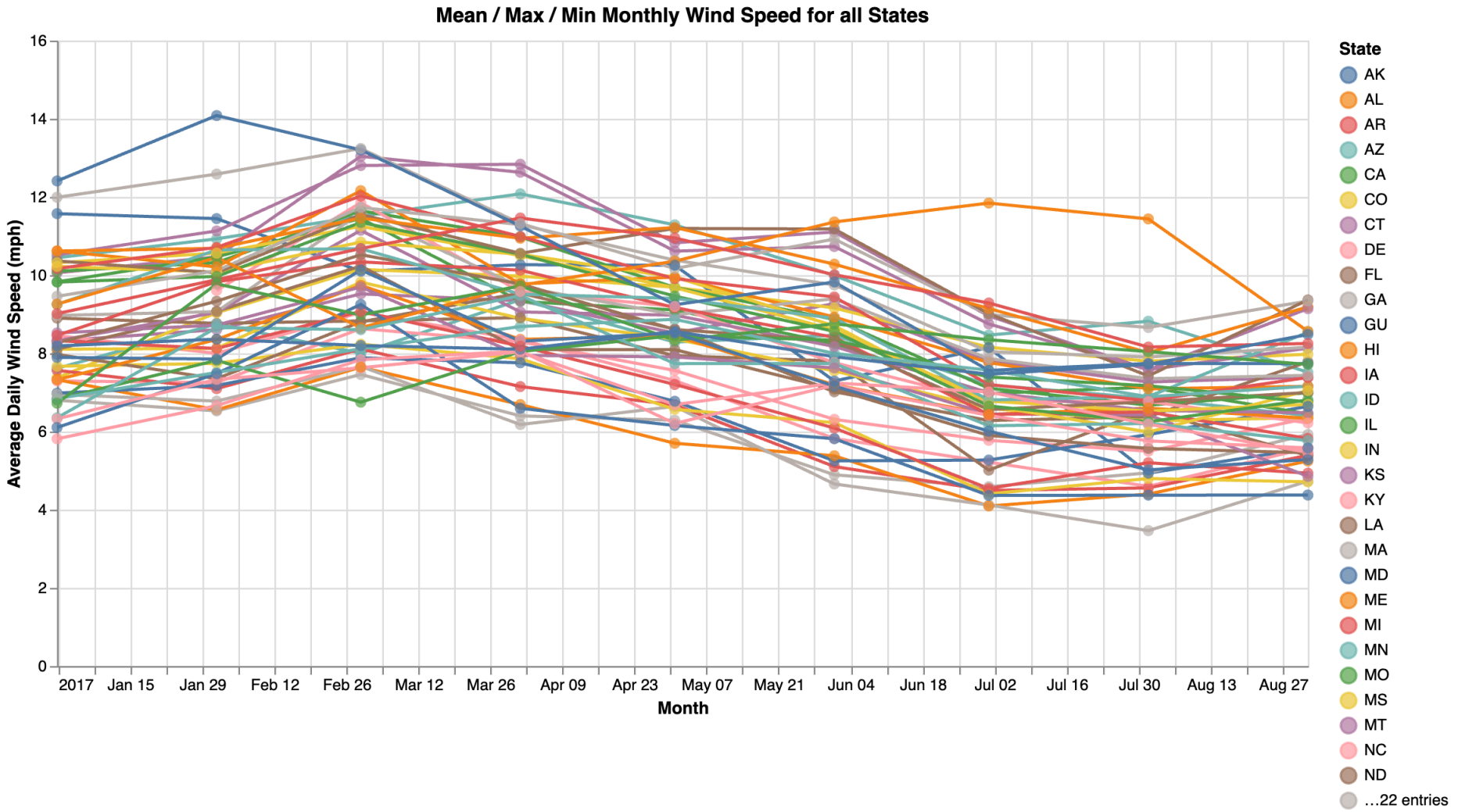


statType mean_PRCP

Monthly Wind Speed

Findings:

- Notable spikes in wind speed appear in TX (spring/summer storms https://www.weather.gov/mob/cindy).
- AK and FL can also see higher wind during hurricane months.

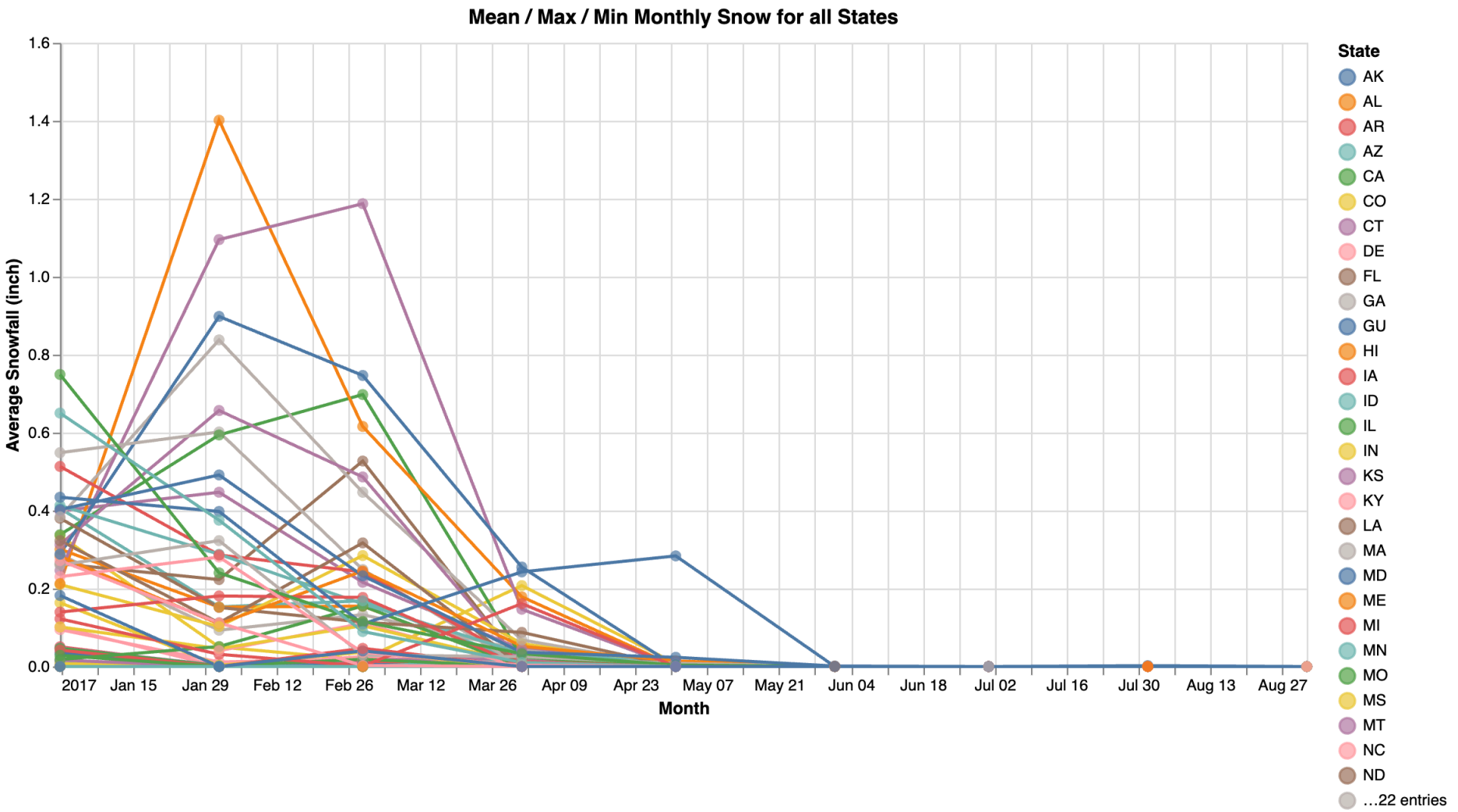


statType mean_AWND

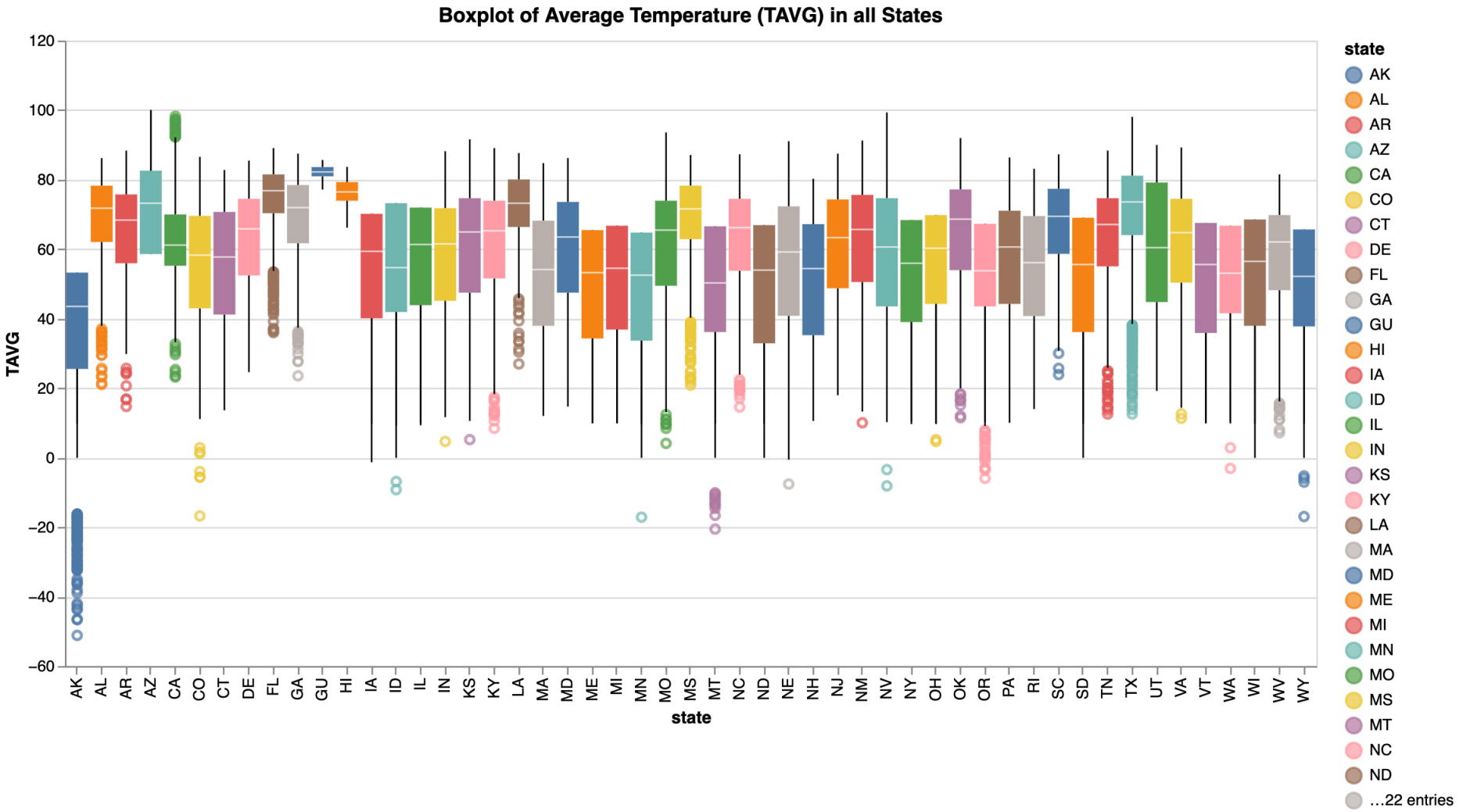
Monthly Snowfall

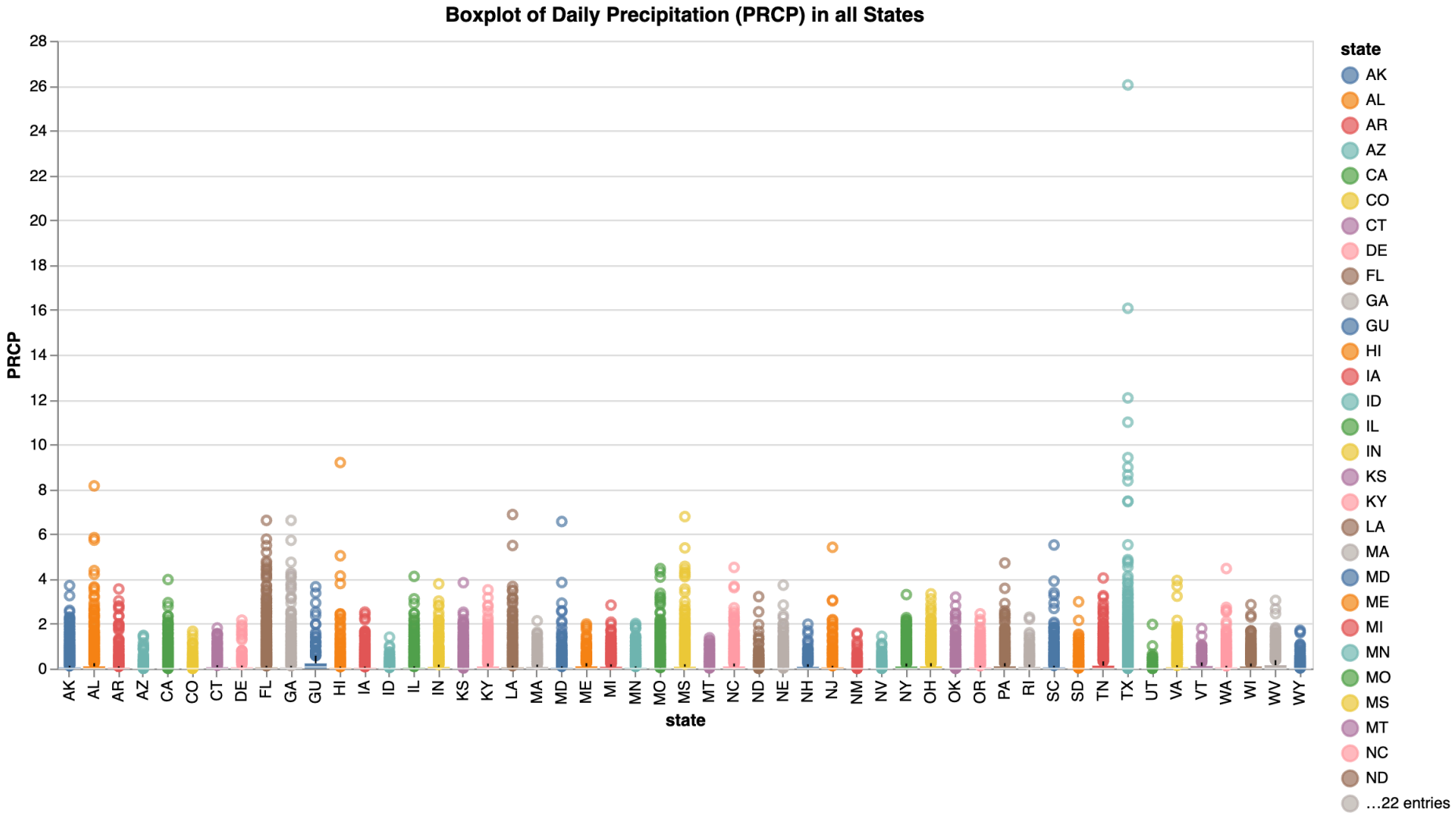
Findings:

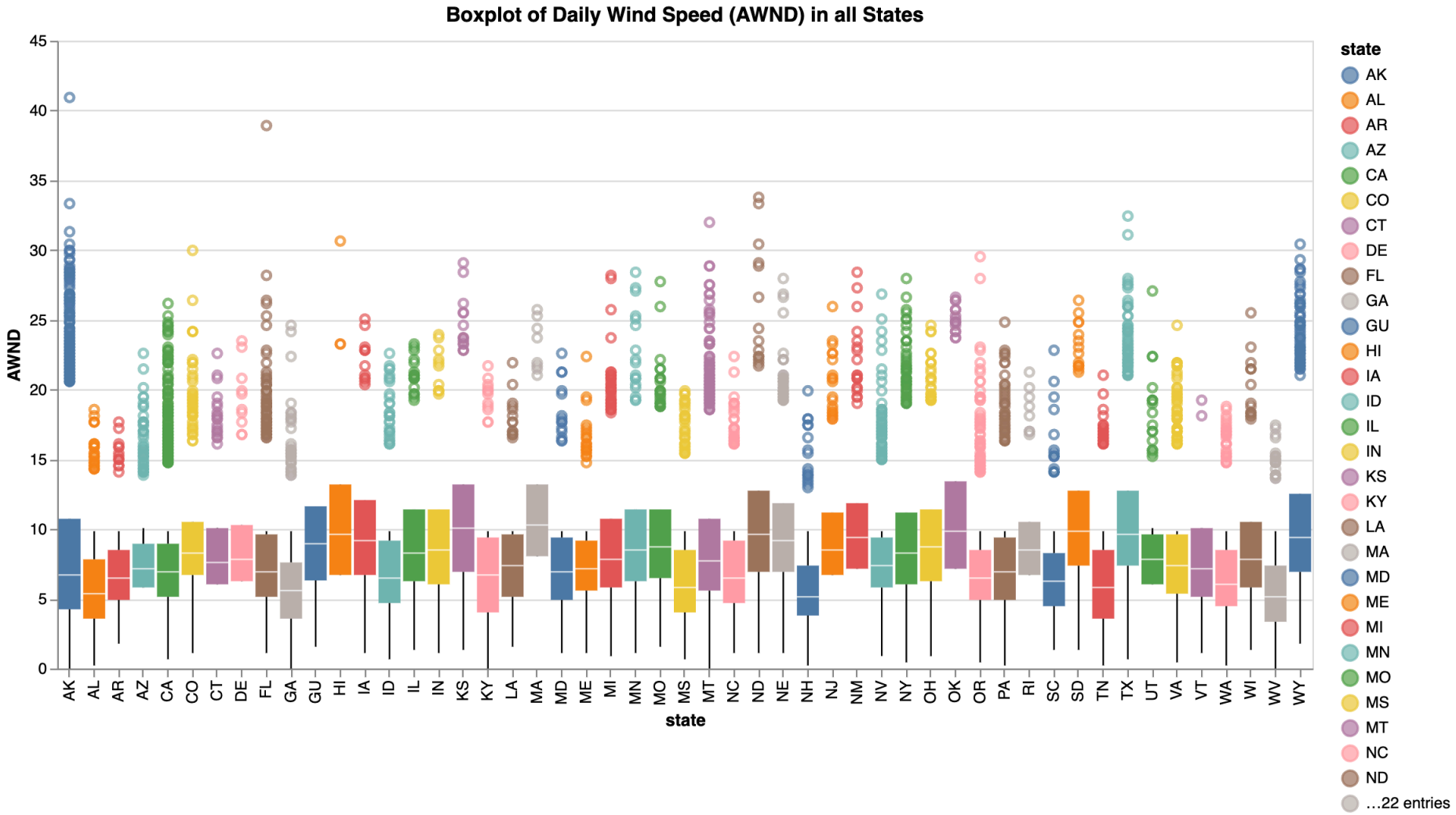
- Substantial snowfall largely restricted to ME from roughly Janurary to March.
- NY experienced a snowstorm between February and March.

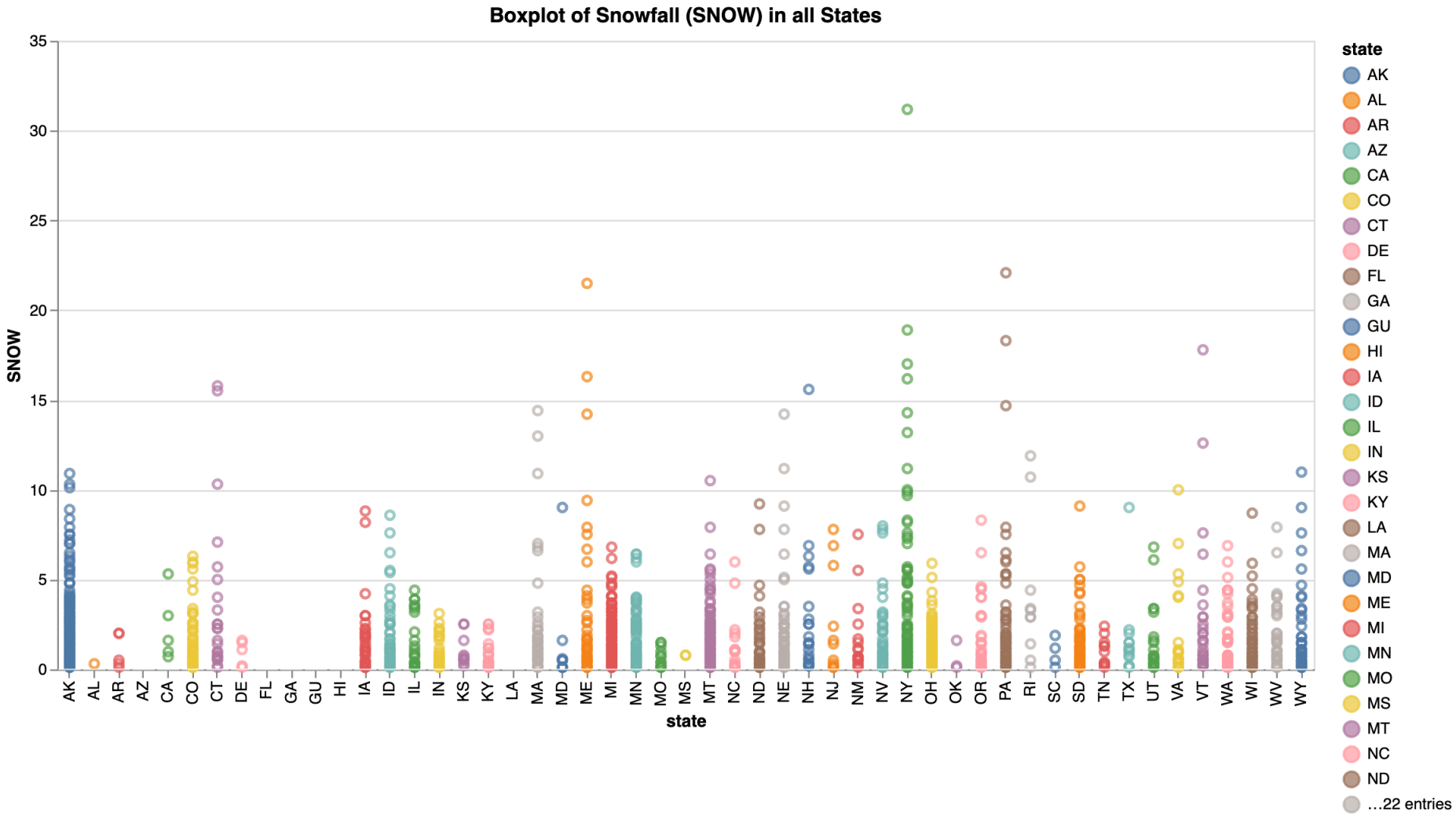


statType mean_SNOW


Boxplot of Average Temperature (TAVG) in all States
TAVG
120
100
80
60
40
20
0
−20
−40
−60
AK AL AR AZ CA CO CT DE FL GA GU HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state
state
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
GU
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
…22 entries


Boxplot of Daily Precipitation (PRCP) in all States
PRCP
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
AK AL AR AZ CA CO CT DE FL GA GU HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state
state
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
GU
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
…22 entries

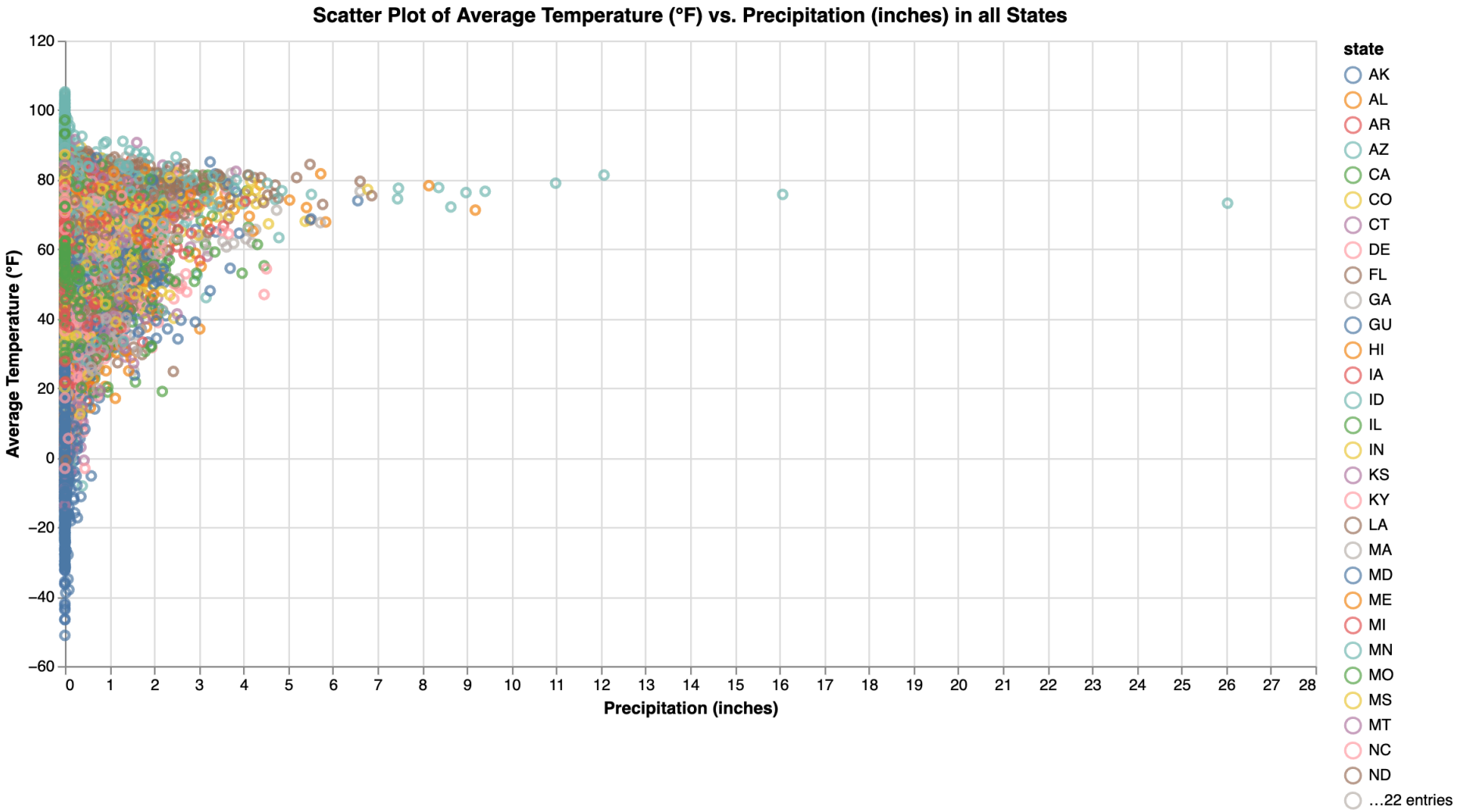

Boxplot of Daily Wind Speed (AWND) in all States
AWND
45
40
35
30
25
20
15
10
5
0
AK AL AR AZ CA CO CT DE FL GA GU HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
state
state
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
GU
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
…22 entries



Scatter Plots and Correlations

TAVG vs. PRCP Observations:

- No strong negative or positive correlation emerges—heavy rain occurs across a wide temperature range. Implication:
  - Precipitation in these states may be more influenced by storms or regional weather patterns rather than strictly temperature.



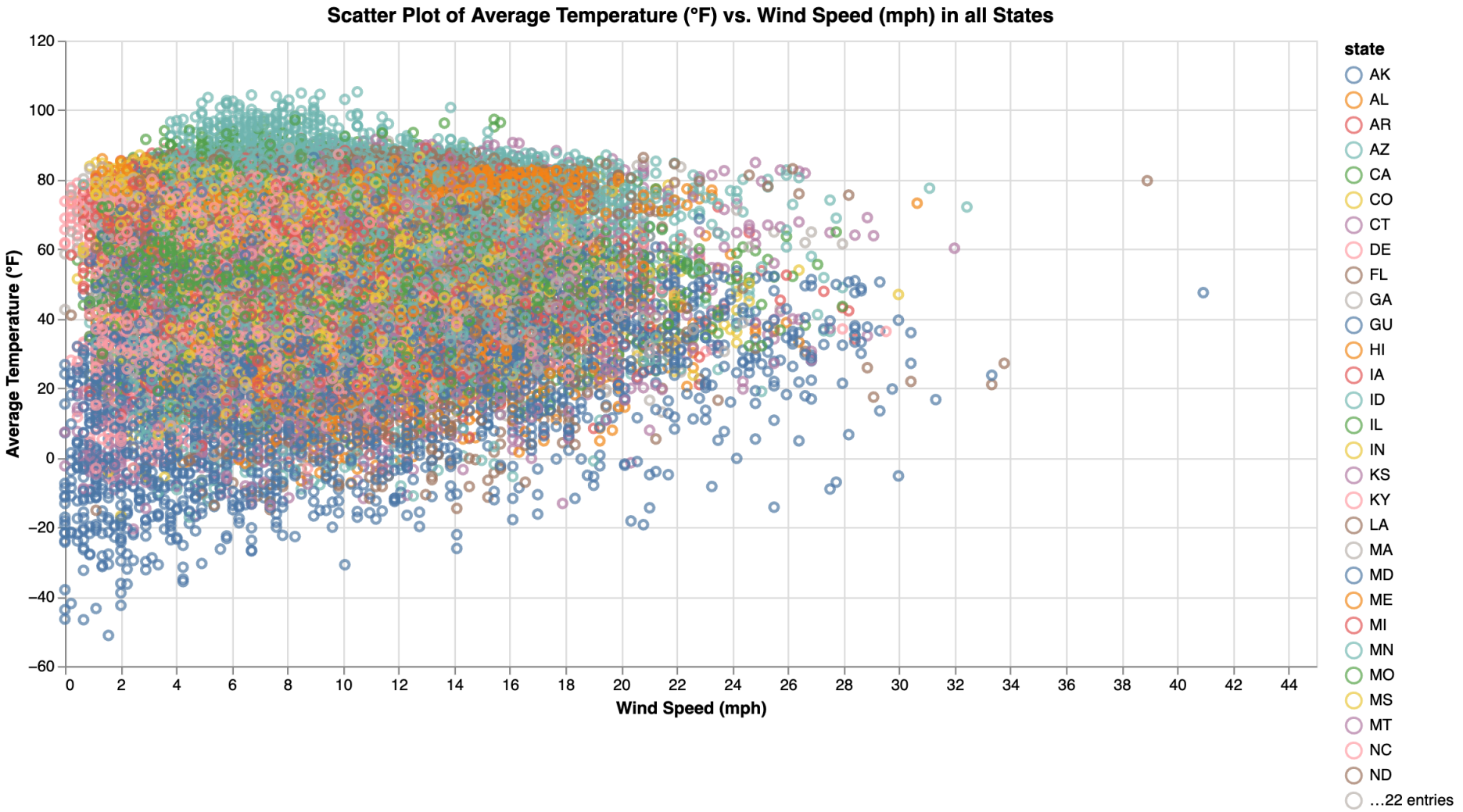
TAVG vs. AWND

Observations:

- Generally, wind speeds cluster around 0–20 mph, but some outliers appear at lower temperatures, hinting at cold fronts or winter storms.

Conclusion:

- Wind speed does not have a simple linear relationship with temperature, but extremes can be tied to specific meteorological events.



states = ▸Array(51) ["AK", "AL", "AR", "AZ", "CA", "CO", "CT", "DE", "FL", "GA", "GU", "HI", "IA", "ID", "IL", "IN", "KS",

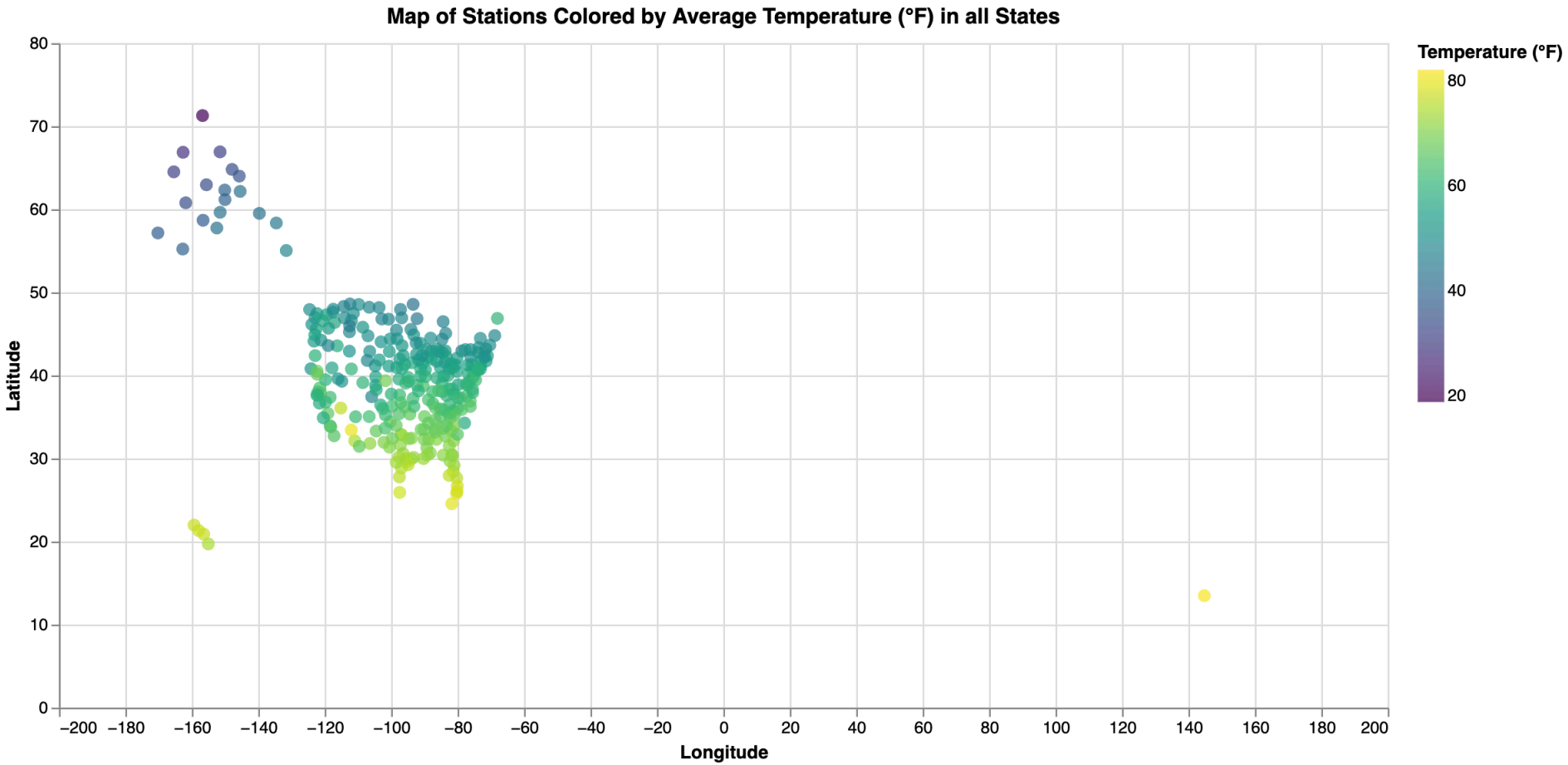
Geographic Scatter (Longitude vs. Latitude, Colored by TAVG)

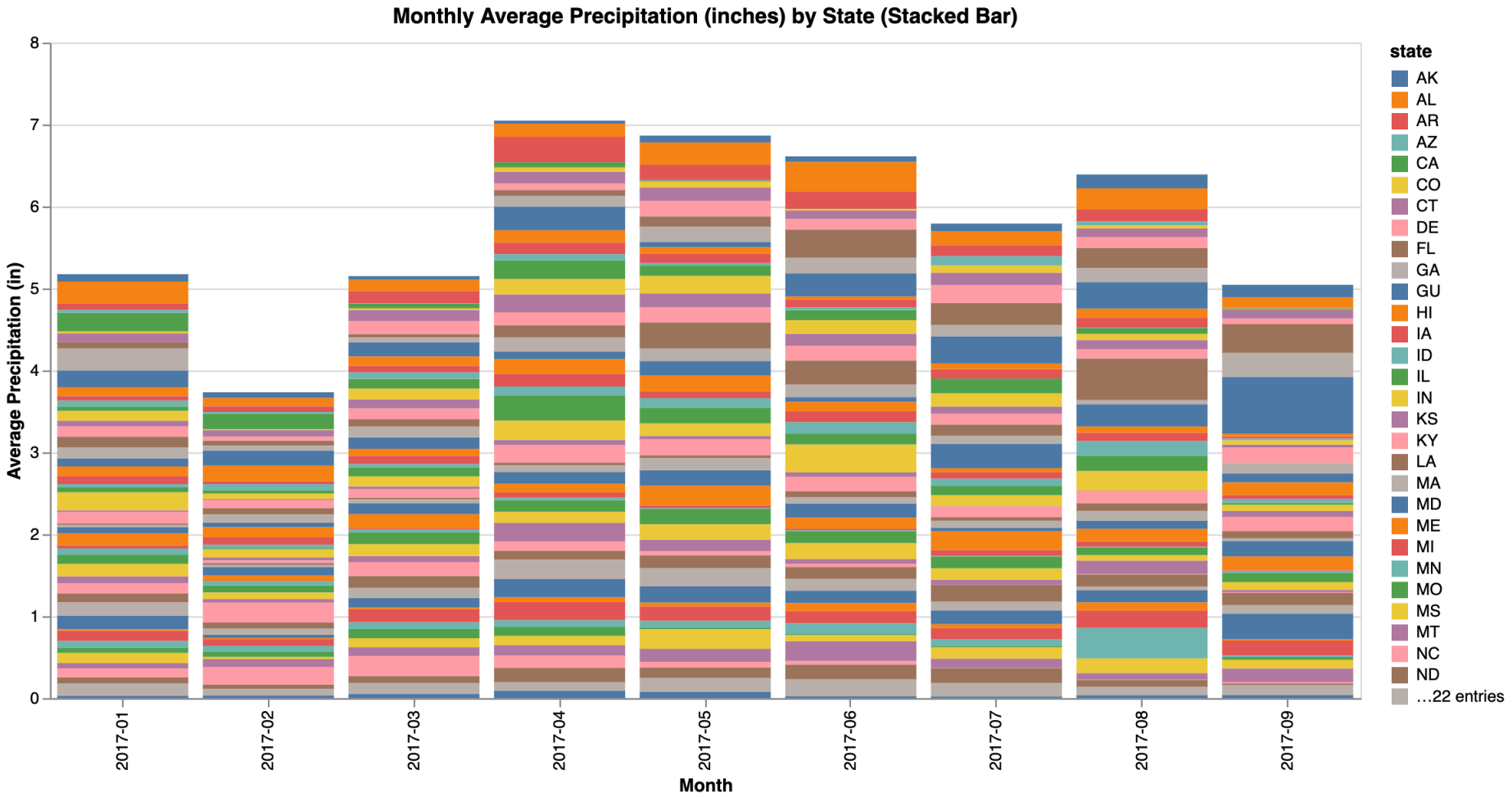
Observations:

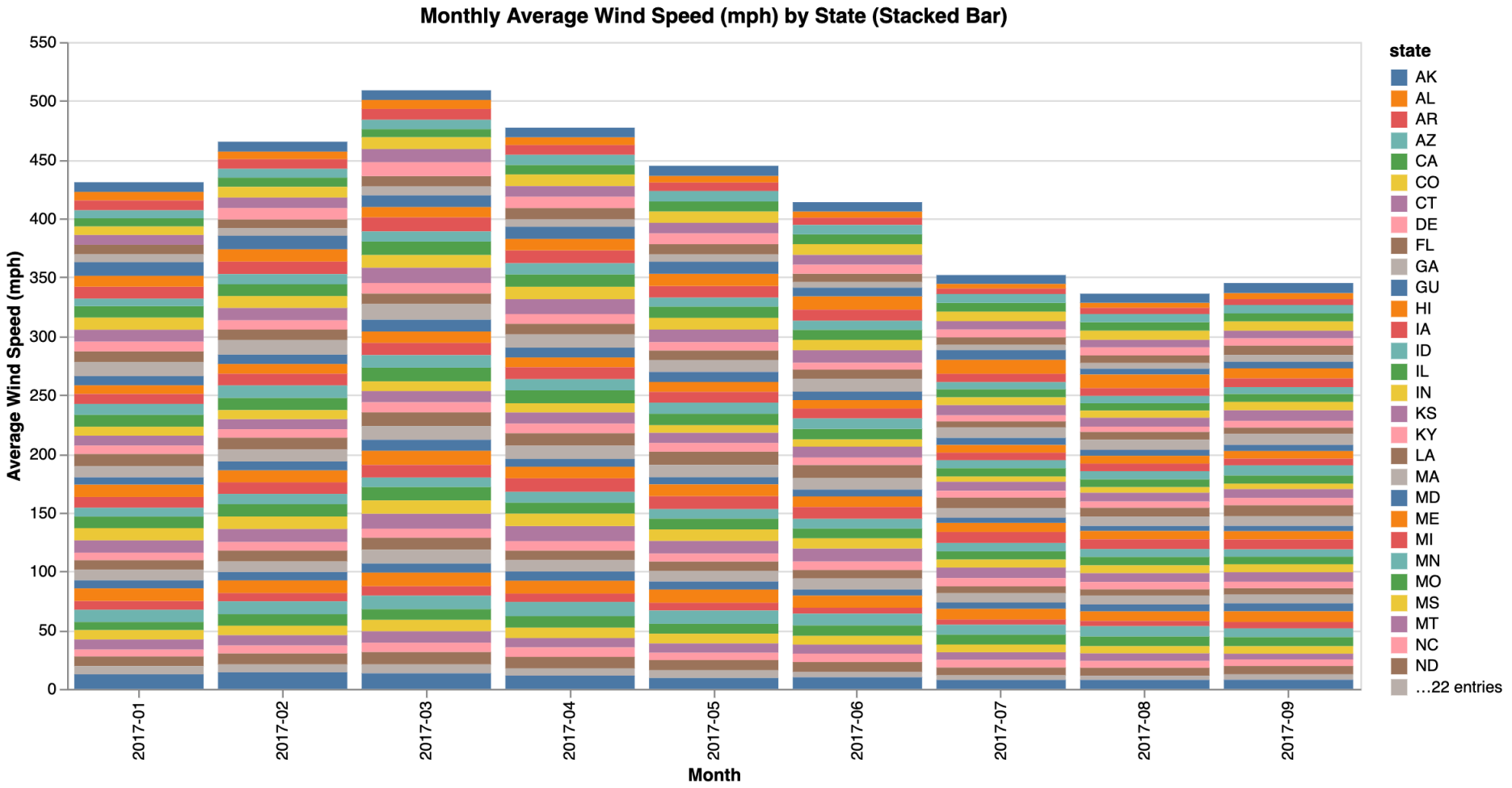
- Warmer colors cluster in the south, cooler in the north.

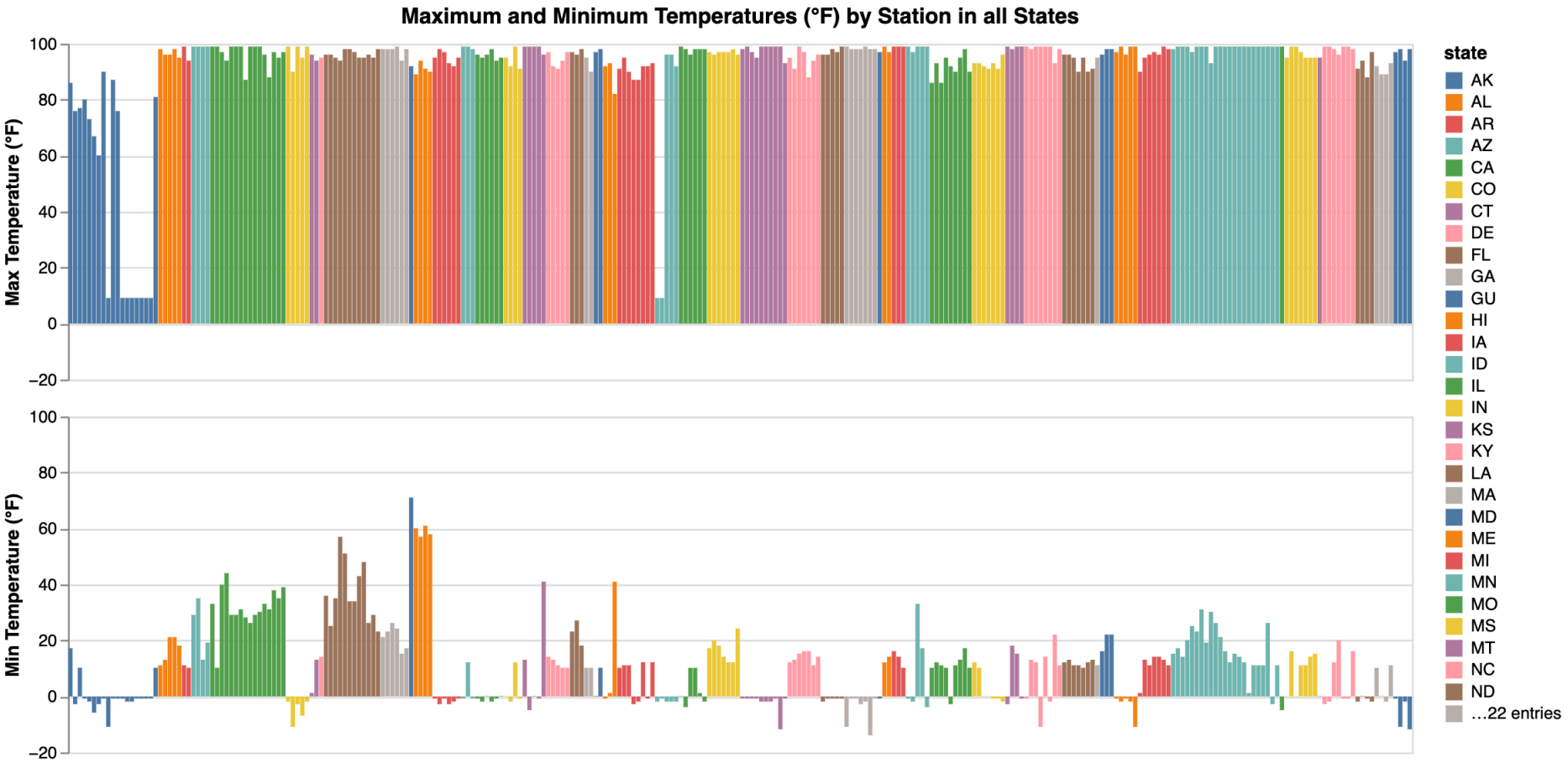
Conclusion:

- A spatial view quickly highlights the latitudinal temperature gradient and helps identify which stations are in warmer or colder zones.


Map of Stations Colored by Average Temperature (°F) in all States
Temperature (°F)
80
60
40
20
Latitude
80
70
60
50
40
30
20
10
0
Longitude
−200 −180 −160 −140 −120 −100 −80 −60 −40 −20 0 20 40 60 80 100 120 140 160 180 200


Monthly Average Precipitation (inches) by State (Stacked Bar)
Average Precipitation (in)
0
1
2
3
4
5
6
7
8
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
Month
state
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
GU
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
…22 entries


Monthly Average Wind Speed (mph) by State (Stacked Bar)
Average Wind Speed (mph)
550
500
450
400
350
300
250
200
150
100
50
0
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
Month
state
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
GU
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
…22 entries


Maximum and Minimum Temperatures (°F) by Station in all States
Max Temperature (°F)
Min Temperature (°F)
100
80
60
40
20
0
−20
state
AK
AL
AR
AZ
CA
CO
CT
DE
FL
GA
GU
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
…22 entries