IMPACT OF WEATHER PATTERNS ON SWAINSON HAWK MIGRATION

By Anthony Wolfe

AGENDA

- Introduction
- Data
- Cleaning & Transformation
- Visualizations & Insights
- Analysis & Conclusion
- References & Citations
- Retrospective

- We're investigating the association and trends between atmospheric variables, such as wind, sun, precipitation, temperature, and Swainson hawk migration patterns on short- and long-term intervals.

- We'd like to know if climate change over the years has somehow changed the behavior of these hawks because of our atmosphere changing with global climate change. Their migration patterns are heavily determined by the condition of the atmosphere, and we should be prepared for any ecological consequences as a result.

- Well, the hawks! In all seriousness though, if we derive some insights from this EDA that tell us that there are some ongoing changes with the behavior of hawks, researches, land conservation groups, ecologists, and more can prepare appropriately and make changes to their approach when considering how the behavior of these raptor birds change with the climate

INTRODUCTION

- What are we doing?
- Why are we doing it?
- Who benefits from this?

DATA

- What data are we using?
- Why are we using these datasets?
- Any considerations worth noting?

- There are three sources we'll be leveraging for our EDA:
  - Global Biodiversity Information Facility (GBIF) – stores animal occurrences across the World from separate publishing organizations
  - Movebank – animal tracking data repository for researchers who'd like to share data they collected or are collecting for previous or live studies
  - National Oceanic & Atmospheric Administration (NOAA) – collects oceanic and atmospheric data from stations around the US and stores the info.

These datasets either provide reliable tidy data or provide a wide range of collected data across different publishing organizations, which allows for diverse analysis across different types of datasets.

- We can make the following considerations while working with these datasets:
  - Consider the fact that we're using occurrence data and tracking data separately against our weather data. We should keep those two separate.
  - Consider the fact that the animal data is either publicly provided or through research studies, which means there can be issues with the data
  - Animal tracking data can be isolated geographically and in terms of time due to many potential reasons

CLEANING AND TRANSFORMATIONS

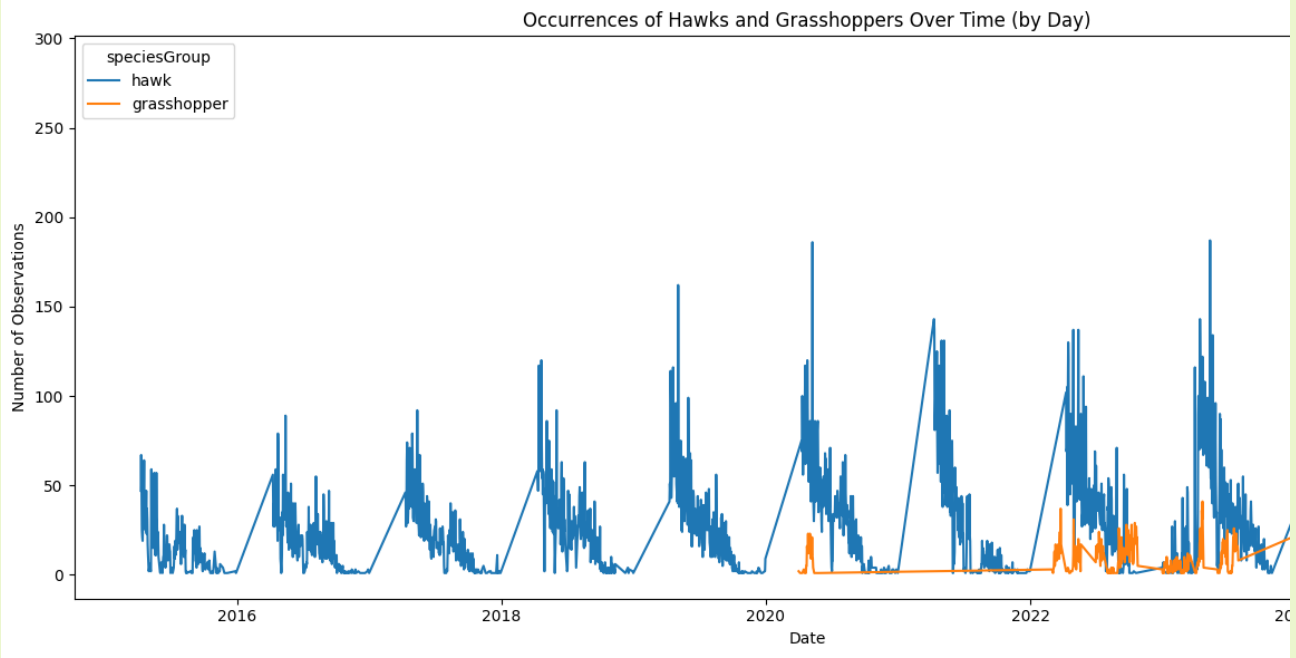
	publishingOrgKey	key	datasetKey	scientificName	decimalLatitude	decimalLongitude	dateIdentified	eventTime	occurrenceStatus	taxonID	taxonRank	parsedDate	speciesGroup
0	28eb1a3f-1c15-4a95-931a-4af90ecb574d	4011915312	50c9509d-22c7-4a22-a47d-8c48425ef4a7	Buteo swainsoni Bonaparte, 1838	33.813019	-116.525763	2023-01-03T01:16:53	10:52:00-08:00	PRESENT	5196	SPECIES	2023-01-02 10:52:00	hawk
1	28eb1a3f-1c15-4a95-931a-4af90ecb574d	4028968442	50c9509d-22c7-4a22-a47d-8c48425ef4a7	Buteo swainsoni Bonaparte, 1838	33.995642	-118.096413	2023-01-27T03:36:07	09:00:00-08:00	PRESENT	5196	SPECIES	2023-01-26 09:00:00	hawk
2	28eb1a3f-1c15-4a95-931a-4af90ecb574d	4028827735	50c9509d-22c7-4a22-a47d-8c48425ef4a7	Buteo swainsoni Bonaparte, 1838	33.995499	-118.096327	2023-01-27T22:22:12	10:58:00-08:00	PRESENT	5196	SPECIES	2023-01-26 10:58:00	hawk

	event-id	visible	timestamp	location-long	location-lat	ground-speed	heading	height-above-ellipsoid	migration-stage	sensor-type	individual-taxon-canonical-name	tag-local-identifier	individual-local-identifier	study-name
0	1583635119	True	2011-06-25 16:00:00.000	-121.62333	38.69267	5.14440	149.0	80.0	nestling	gps	Buteo swainsoni	105921	105921	Space use by Swainson's Hawk (Buteo swainsoni)...
1	1583635120	True	2011-06-25 17:00:00.000	-121.62400	38.69433	5.14440	150.0	160.0	nestling	gps	Buteo swainsoni	105921	105921	Space use by Swainson's Hawk (Buteo swainsoni)...
2	1583635121	True	2011-06-25 18:00:00.000	-121.62500	38.69383	5.14440	226.0	190.0	nestling	gps	Buteo swainsoni	105921	105921	Space use by Swainson's Hawk (Buteo swainsoni)...

	date	station	latitude	longitude	state	name	awnd	prcp	tavg	tsun
0	2016-01-01	US1CAAL0001	37.812	-122.21	0.4	CA,PIEDMONT 1.0 SE	5.601335	0.0	42.869616	0.0
1	2016-01-01	US1CAAL0003	37.716	-122.05	0.4	CA,CASTRO VALLEY 0.4 NNE	5.601335	0.0	42.869616	0.0
2	2016-01-01	US1CAAL0004	37.648	-121.87	0.0	CA,PLEASANTON 1.8 SSE	5.601335	0.0	42.869616	0.0

- GBIF Summary: Values were in the wrong column in some cases such as date, latitude, and longitude. Some studies collected from GBIF used columns more than others, so there were some null values which needed removing or filling
- Movebank Summary: Only one study contained relevant data, so it was mostly clean, however we needed to still needed to drop some rows with null values, fill others if most rows contained null values, and remove redundant columns.
- NOAA Summary: Data is already tidy and follows a format so no worries there. Only issues were pairing the station lat/lng pairs via another csv and capturing a complete time series since missing data is almost guaranteed. Also, we split this into two data-frames by time ranges corresponding to the GBIF and Movebank time ranges.

VISUALIZATIONS AND INSIGHTS (1)



The line plot below displays the number of observations made in the GBIF dataset over time, which helps us see some potential gaps in the dataset. We also see it follow a seasonal trend, creating new heights each year. Blue are the hawk occurrences and orange are the grasshoppers.

VISUALIZATIONS AND INSIGHTS (2)

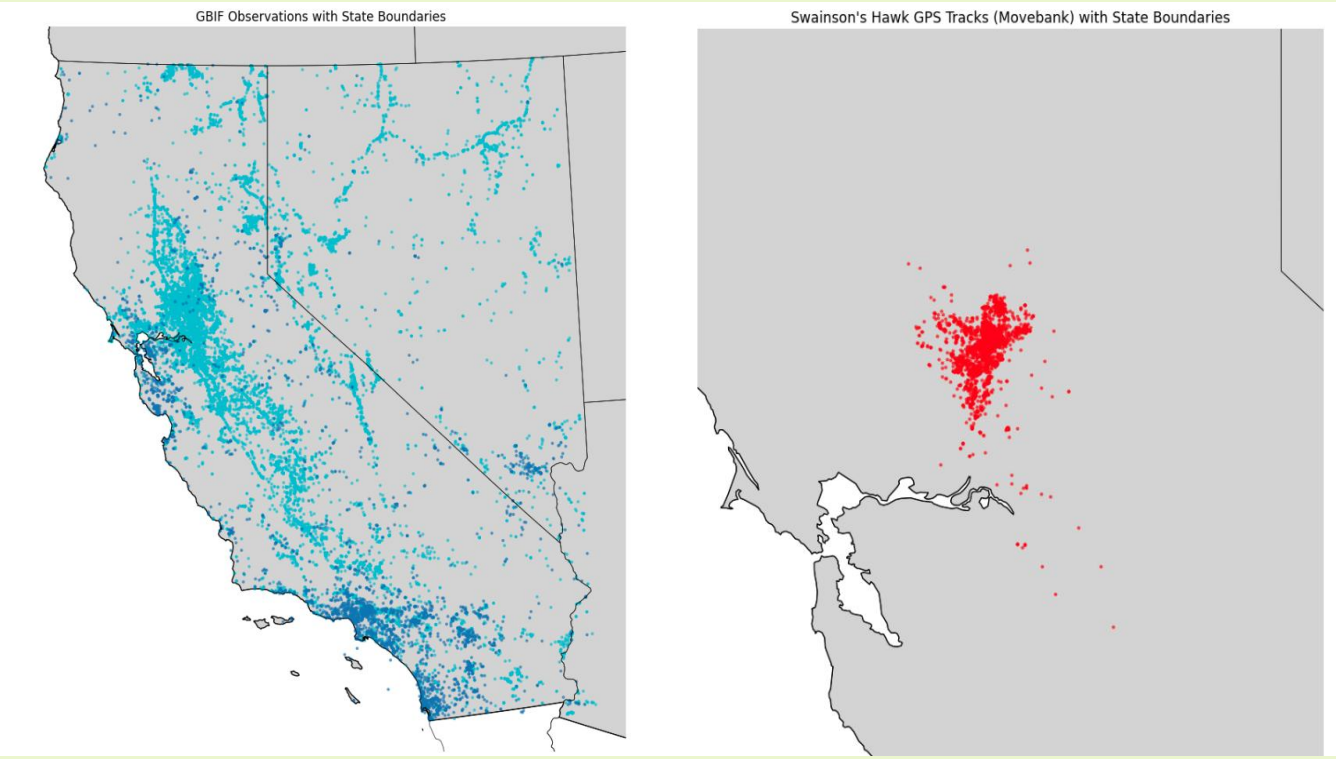
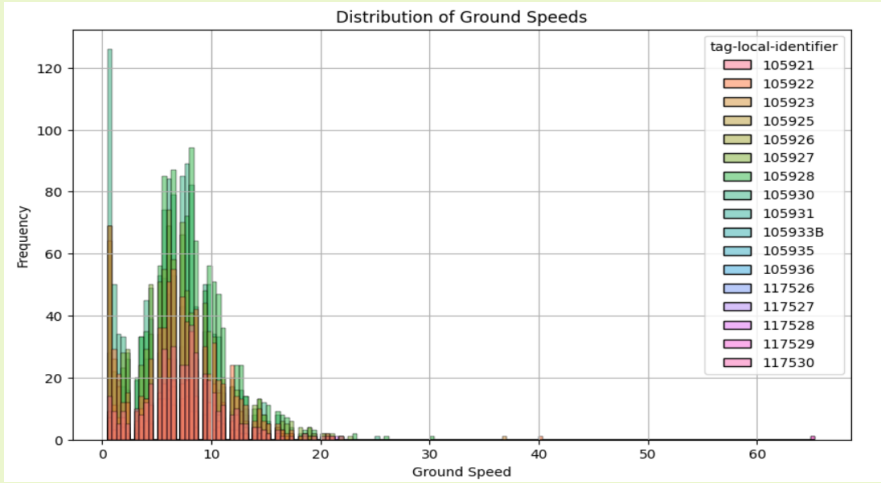
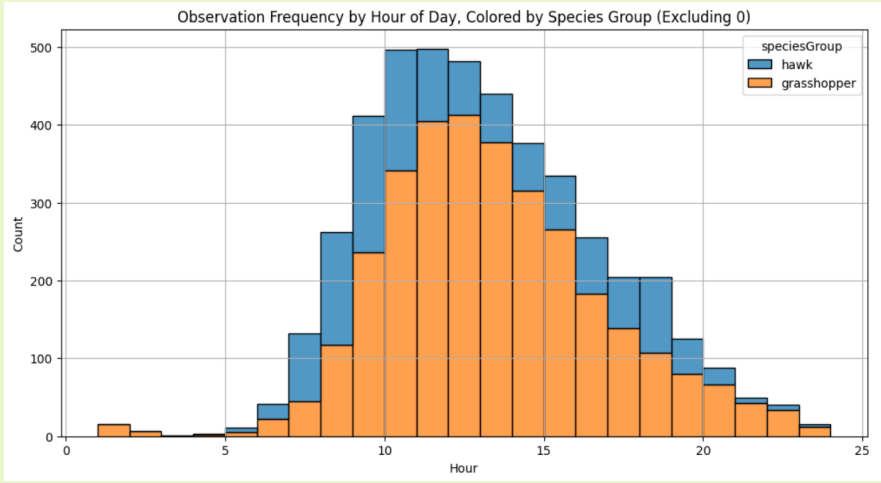


Image on the left shows occurrences across 2015 through 2025 for Swainson hawks in light blue and grasshoppers in blue on a geo scatter plot.

Image on the right is a geo scatter plot that shows all hawk related tracking records from Movebank between 2011 and 2013, which doesn't capture each individual subject.

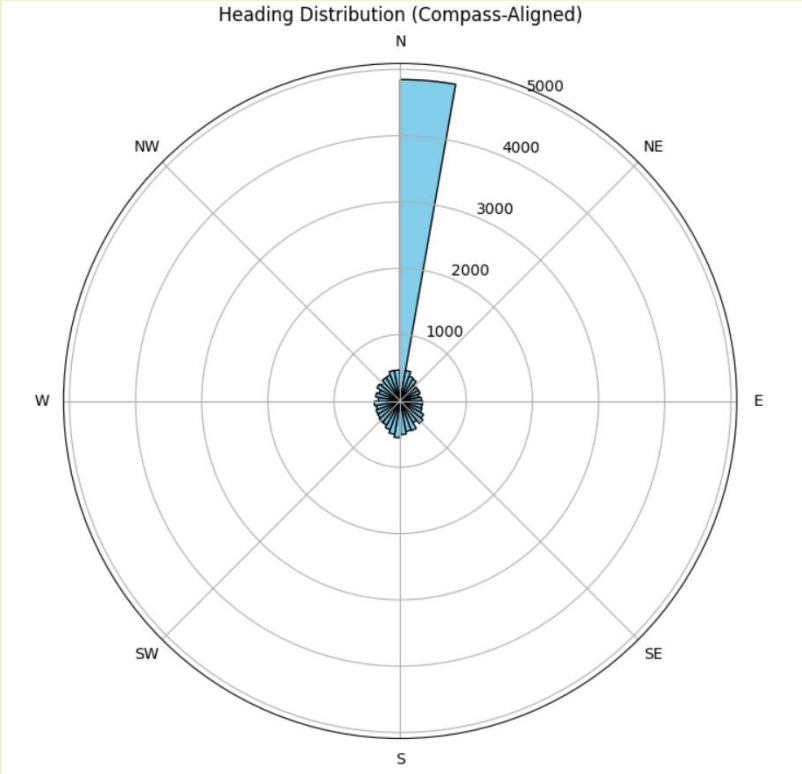
VISUALIZATIONS AND INSIGHTS (3)





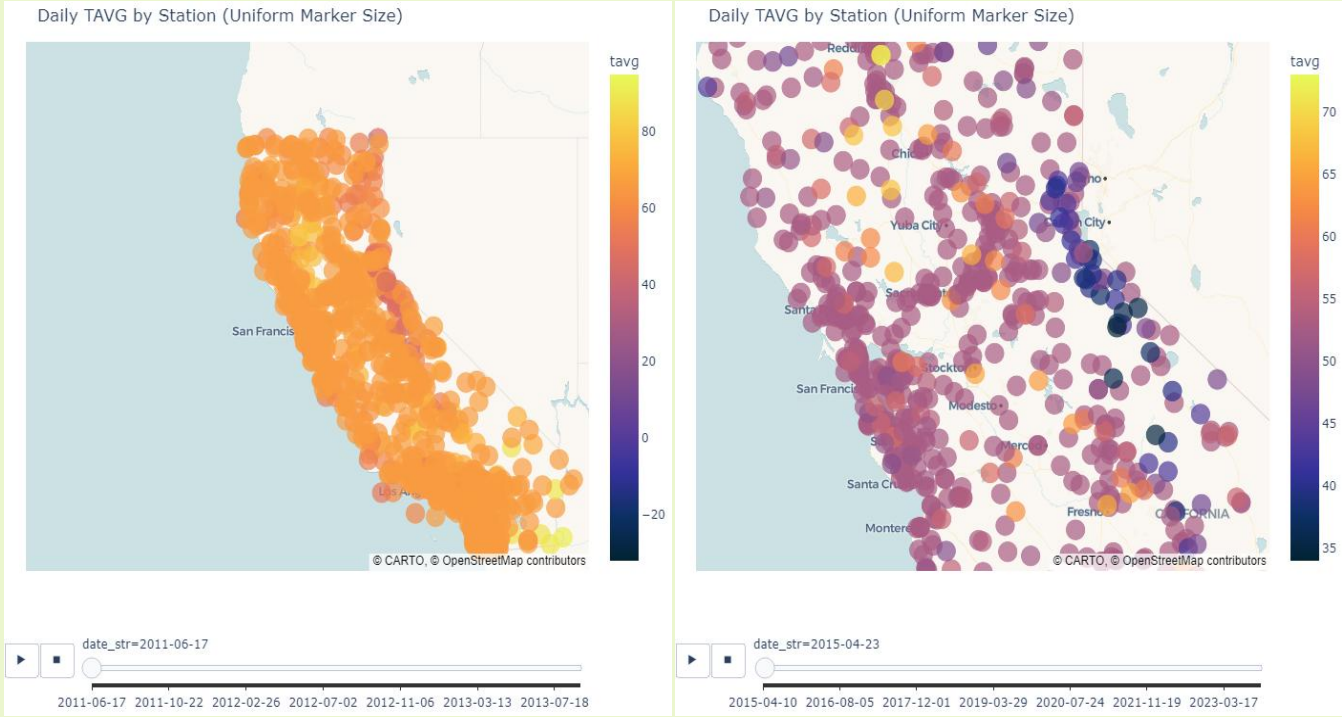
The plot above visualizes a count plot distribution of ground speeds captured in the Movebank dataset for hawks, hued by each individual subject. Something interesting is that they all follow a similar distribution of speeds. The one below shows the count plot distribution of GBIF occurrences based on the time they were recorded, hued by either hawk or grasshopper.

VISUALIZATIONS AND INSIGHTS (4)



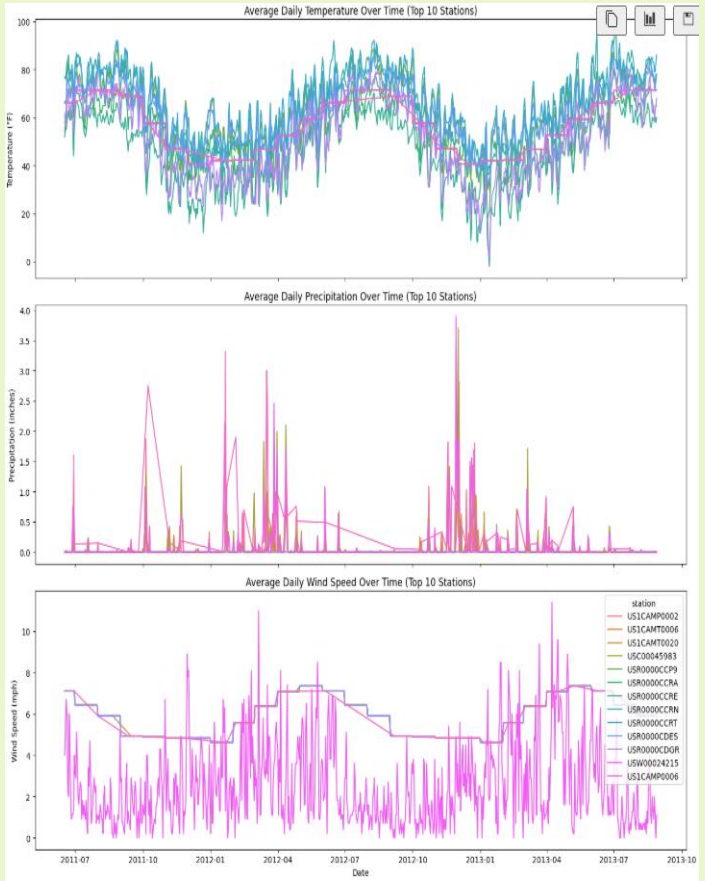
The visualization above captures a distribution of headings, which is a value representing the angle where the hawk was flying, somehow determined by the GPS. I appropriately angled and labeled it so you could see the distribution of where most hawks are flying within the Movebank dataset where individual hawks were being tracked for a separate study.

VISUALIZATIONS AND INSIGHTS (5)



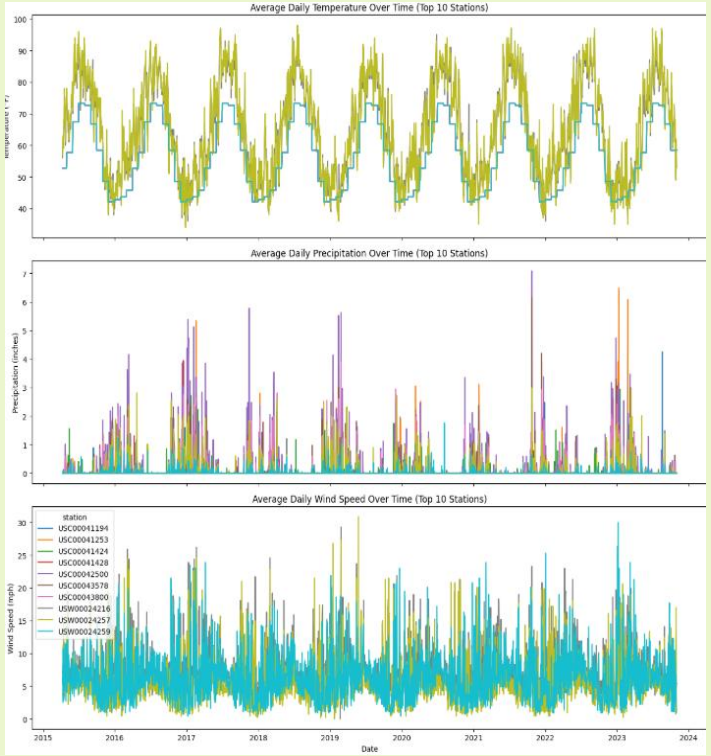
The above static image is a plotly-generated scatter-map, which displays the values for certain weather variables overtime using a heatmap circle for the given station's latitude-longitude pairs. The one on the left was meant for the Movebank and NOAA merged dataset, however on the right is GBIF and NOAA's merged dataset

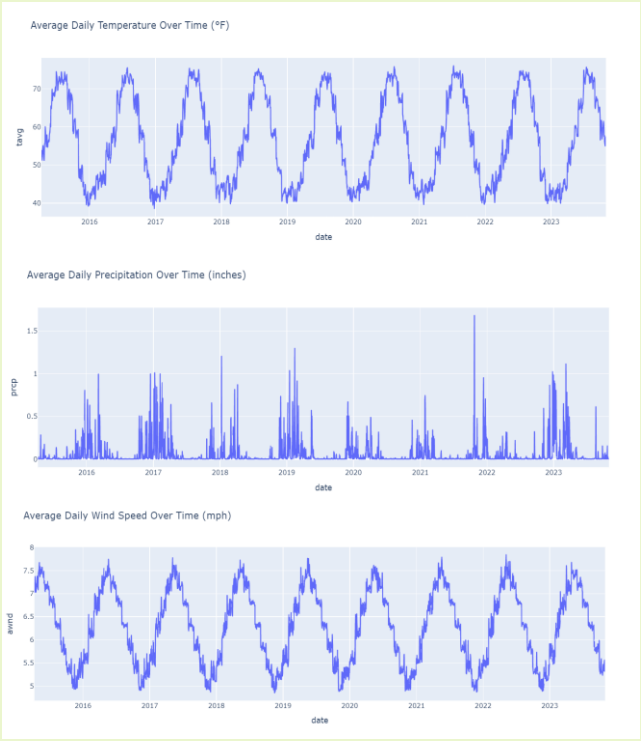
VISUALIZATIONS AND INSIGHTS (6)



The three plots on the right depict the average weather values (temp, wind, rain) across all these CA stations relevant to the Movebank timeframe (2011-2013). On the left, it's the same however we hue by the top ten stations in the dataset.

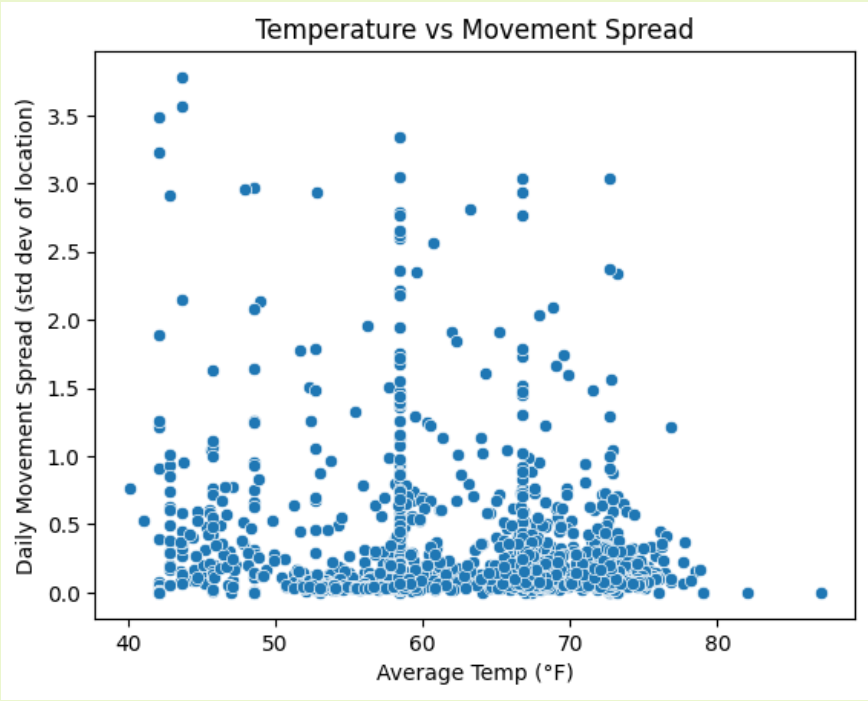
VISUALIZATIONS AND INSIGHTS (7)





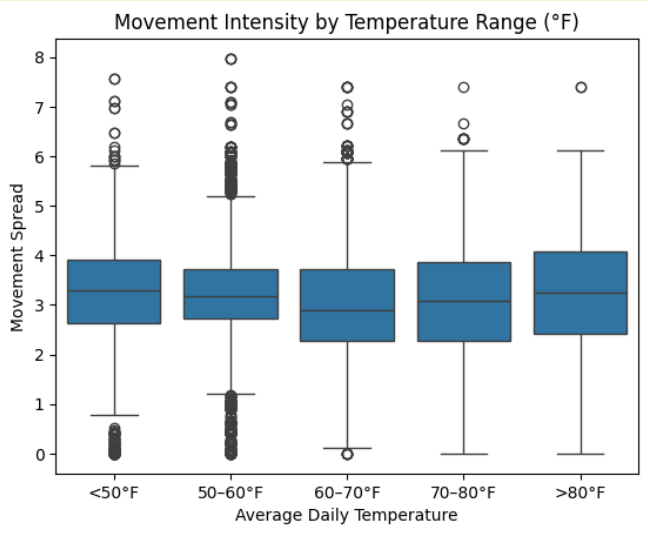
The three plots on the right depict the average weather values (temp, wind, rain) across all these CA stations relevant to the GBIF timeframe (2015-2024). On the left, it's the same however we hue by the top ten stations in the dataset.

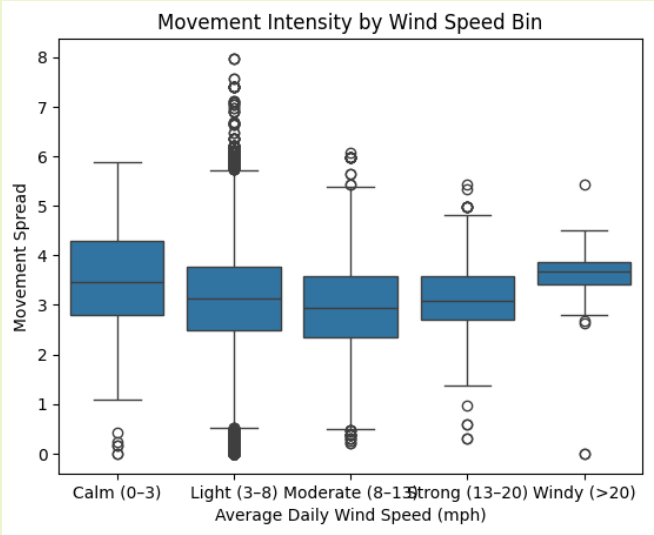
ANALYSIS & CONCLUSIONS (Q1)



Weather Variable	Spearman r	p-value	Interpretation
`tavg` (Temperature)	0.080	0.0006	Very weak positive correlation — warmer days slightly increase movement
`awnd` (Wind Speed)	-0.482	< 0.0001	Moderate to strong negative correlation — high winds suppress movement
`prcp` (Precipitation)	-0.091	0.0001	Very weak negative correlation — rain slightly inhibits movement

ANALYSIS & CONCLUSIONS (Q2)





The permutation test for the left plot above checks the significance between means of the 50-60 Fahrenheit bucket and all the others, which gave us an approximately 0.0001 p-value implying significance and optimal weather for launching or completing migration segments. The one on the right underwent a similar permutation test for the light wind (3-8mph) group, which also gave a similar p-value, indicating potential ideal soaring conditions or transitional migration conditions

RETROSPECTIVE

Advantages

Obstacles

Future Improvements

1. Learned how to explore, clean, and transform ecological data via APIs, python functions, and my own domain knowledge about the weather, California, and animals.

2. Time management, knowledge on geospatial analysis, underestimating data collection and cleaning effort required to complete the assignment.

3. Complete geospatial time series analysis on GBIF, Movebank and NOAA datasets, conclude findings, clean up notebook, publish to GitHub.

REFERENCES AND CITATIONS

1. https://www.ncei.noaa.gov/
2. https://www.gbif.org
   - iNaturalist Research Observations, Birda Global Observational Dataset, EBird Observational Dataset, Observation.org Global Nature Data, Xeno-canto Global Bird Sounds
3. https://www.movebank.org/
   - https://datarepository.movebank.org/entities/datapackage/fb9f260b-b3fa-4c3b-a059-847341c43998

THANK YOU!

Anthony Wolfe | toritotony.github.io