

# Hurricane Data Analysis and Visualization

```
In [2]: import pandas as pd
```

```
In [3]: import altair as alt
alt.renderers.enable('notebook')
```

```
Out[3]: RendererRegistry.enable('notebook')
```

## read in dataset (documented [here](#))

```
In [35]: df = pd.read_csv('hurdat2.csv')
df
```

	identifier	name	num_pts	record_id	status	latitude	longitude	max_wind	min_pressure	datetime
0	AL011851	UNNAMED	14	NaN	HU	28.0	-94.8	80	-999	1851-06-25T00:00:00
1	AL011851	UNNAMED	14	NaN	HU	28.0	-95.4	80	-999	1851-06-25T06:00:00
2	AL011851	UNNAMED	14	NaN	HU	28.0	-96.0	80	-999	1851-06-25T12:00:00
3	AL011851	UNNAMED	14	NaN	HU	28.1	-96.5	80	-999	1851-06-25T18:00:00
4	AL011851	UNNAMED	14	L	HU	28.2	-96.8	80	-999	1851-06-25T21:00:00
...	...	...	...	...	...	...	...	...	...	...
50298	AL192017	RINA	21	NaN	TS	38.3	-48.8	45	994	2017-11-08T12:00:00
50299	AL192017	RINA	21	NaN	TS	40.1	-49.0	45	992	2017-11-08T18:00:00
50300	AL192017	RINA	21	NaN	TS	41.8	-48.8	45	991	2017-11-09T00:00:00
50301	AL192017	RINA	21	NaN	LO	43.6	-48.0	40	993	2017-11-09T06:00:00
50302	AL192017	RINA	21	NaN	LO	45.5	-47.0	40	995	2017-11-09T12:00:00

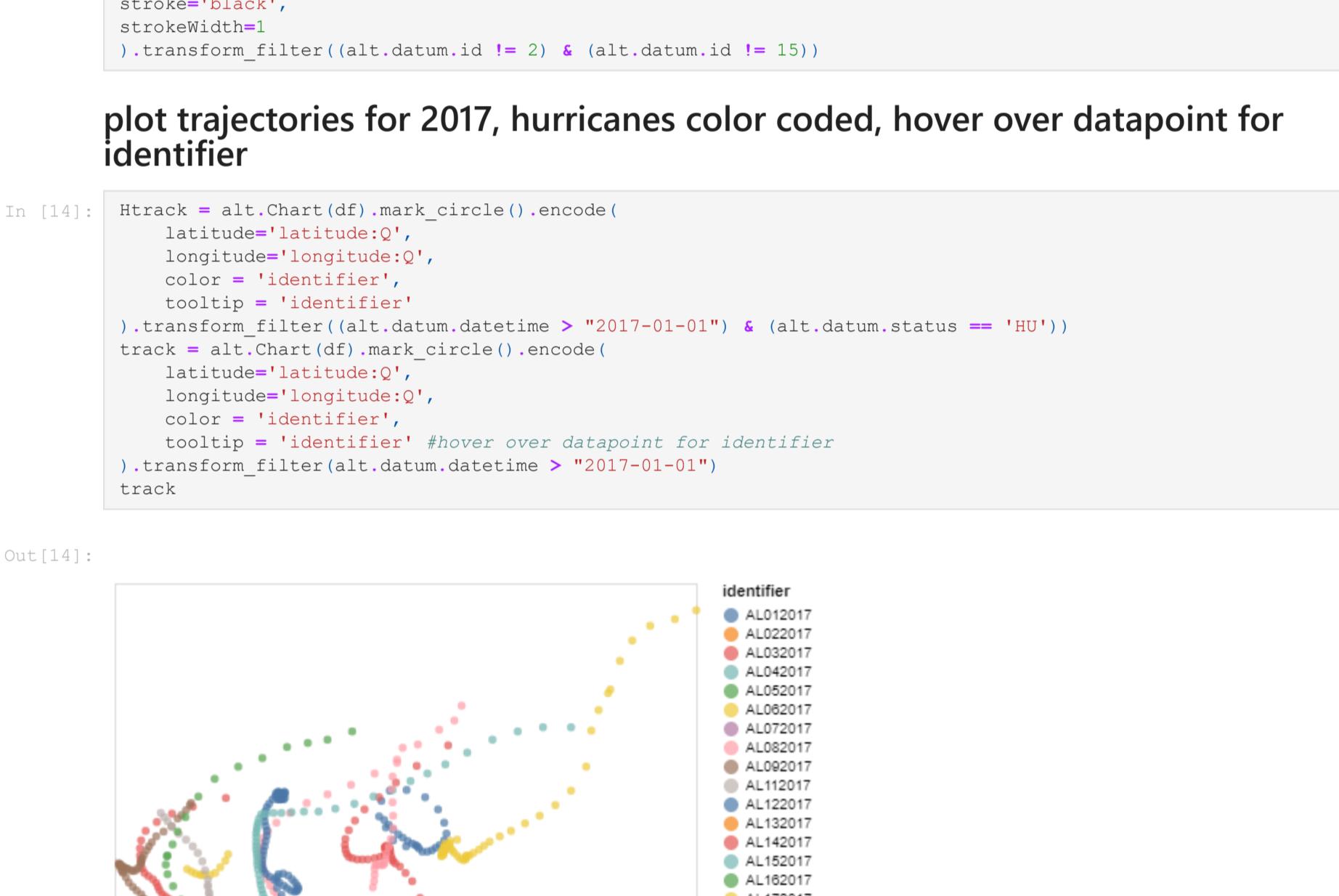
50303 rows × 10 columns

## Bar Chart -- Number of Hurricanes per Year Color Coded by Month

```
In [9]: alt.data_transformers.disable_max_rows()
```

```
Out[9]: DataTransformerRegistry.enable('default')
```

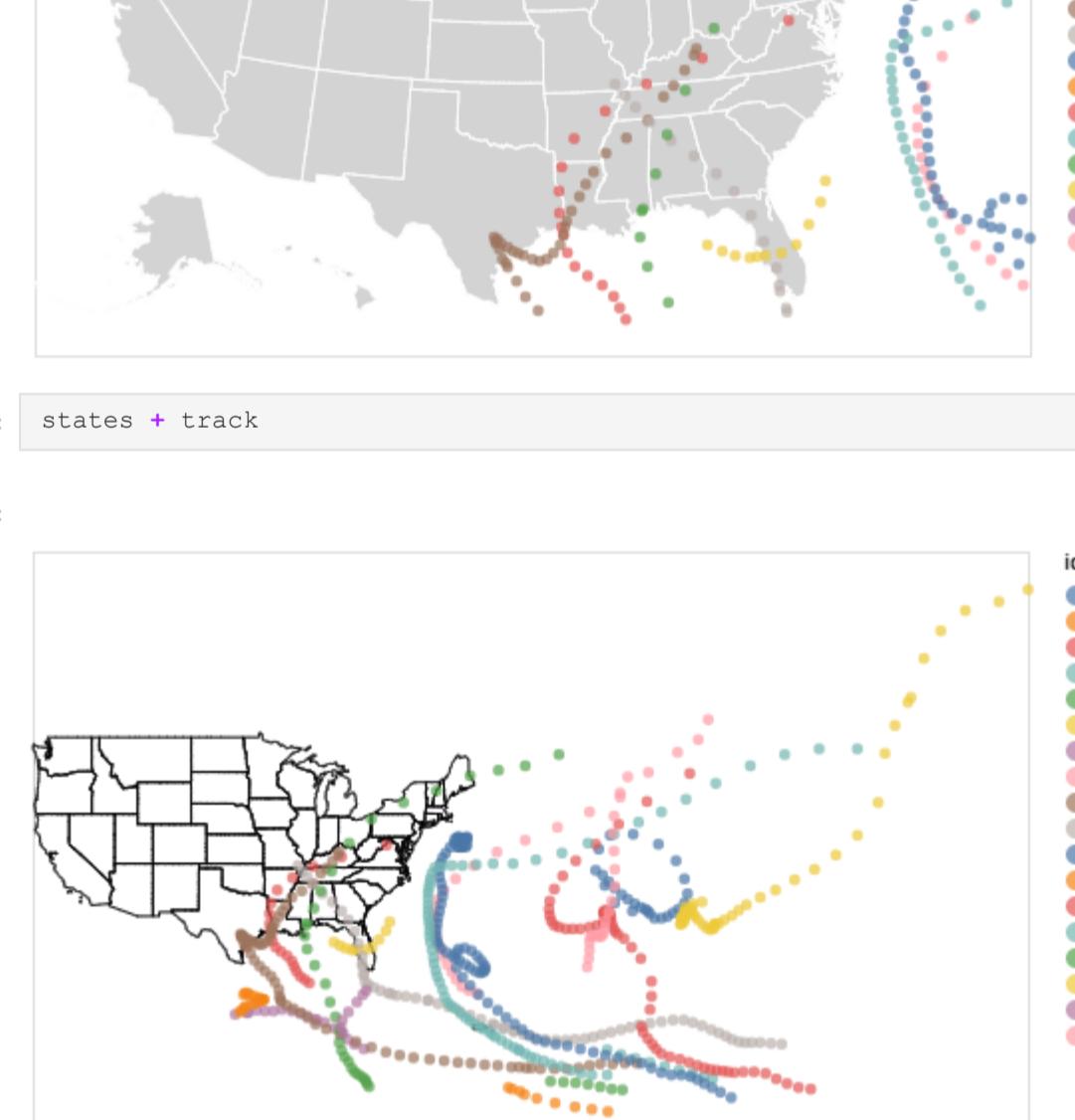
```
In [10]: chart = alt.Chart(df).mark_bar().encode(
    alt.X('year(datetime):O', title='YEAR', scale=alt.Scale(rangeStep=5.5)),
    alt.Y('identifier:O', aggregate='distinct', title='NUMBER OF HURRICANES'),
    color=alt.Color('utcmonth(datetime):O', title='MONTH', scale=alt.Scale(scheme='redyellowblue'))
)
```



## Scatterplot of Wind and Pressure

```
In [11]: alt.Chart(df).mark_circle().encode(
    alt.X('max_wind', scale=alt.Scale(zero=False)),
    alt.Y('min_pressure', scale=alt.Scale(domain=(900, 1025))),
    color=alt.Color('min_pressure', scale=alt.Scale(scheme='viridis')),
    tooltip=['identifier', 'max_wind', 'min_pressure']
).transform_filter(
    (alt.datum.min_pressure != -999)
)
```

```
Out[11]:
```



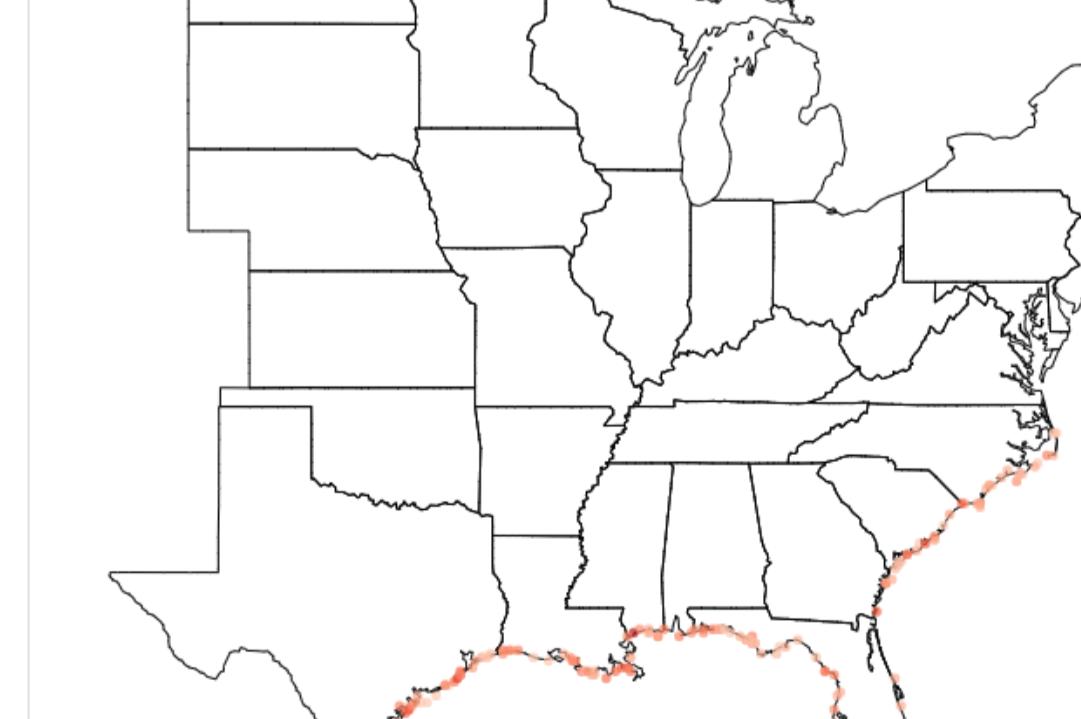
## Hurricane Trajectories

```
In [12]: # get states map
import vega_datasets
from vega_datasets import data
states_data = alt.topo_feature(vega_datasets.data.us_10m.url,
"states")
states = alt.Chart(states_data, width=500, height=300).mark_geoshape(
stroke='black',
strokeWidth=1
).transform_filter((alt.datum.id != 2) & (alt.datum.id != 15))
```

## plot trajectories for 2017, hurricanes color coded, hover over datapoint for identifier

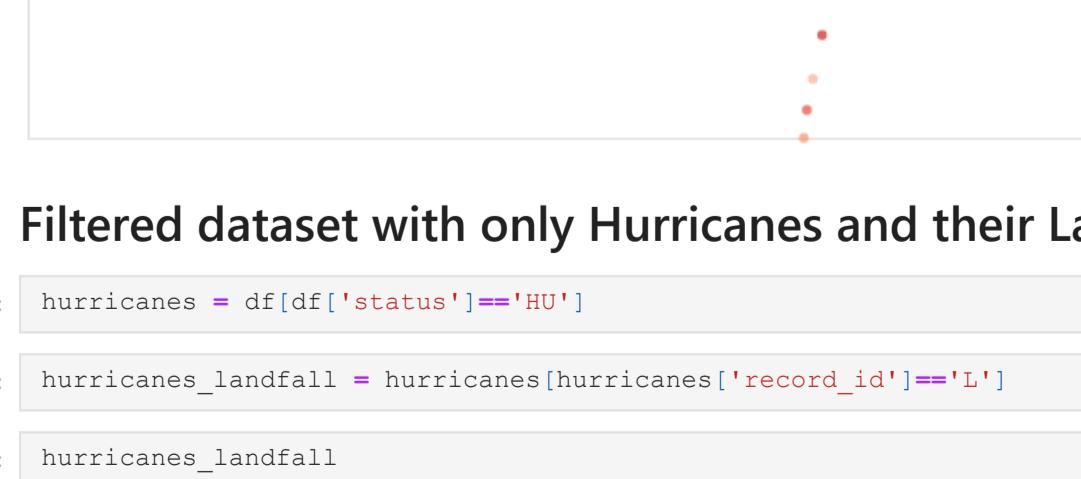
```
In [14]: htrack = alt.Chart(df).mark_circle().encode(
    latitude='latitude:Q',
    longitude='longitude:Q',
    color='identifier',
    tooltip='identifier'
).transform_filter((alt.datum.datetime > "2017-01-01") & (alt.datum.status == 'HU'))
track = alt.Chart(df).mark_circle().encode(
    latitude='latitude:Q',
    longitude='longitude:Q',
    color='identifier',
    tooltip='identifier' #hover over datapoint for identifier
).transform_filter(alt.datum.datetime > "2017-01-01")
track
```

```
Out[14]:
```



```
In [16]: states + track
```

```
Out[16]:
```



## Scatterplot of Max Wind Speed Upon Landfall by Year

```
In [19]: chart = alt.Chart(df).mark_circle().encode(
    alt.X('year(datetime):O', title='YEAR', scale=alt.Scale(rangeStep=5.3)),
    alt.Y('max_wind', title='MAX WIND SPEED', scale=alt.Scale(domain=(60, 170))),
    color=alt.Color('max_wind', scale=alt.Scale(scheme='viridis')),
    tooltip='identifier'
).transform_filter((alt.datum.status == 'HU') & (alt.datum.record_id == 'L')) #gets only hurricanes at landfall
```

```
Out[19]:
```



## Plot Hurricanes' Point of Landfall and Max Wind Speed at Landfall

```
In [38]: # get states map and edit states showing
states2 = alt.Chart(states_data).mark_geoshape(
stroke='black',
strokeWidth=1
).transform_filter((alt.datum.id != 2) & (alt.datum.max_wind != 15) & (alt.datum.id != 41) & (alt.datum.id != 53) &
```

```
In [39]: # plot point where hurricane hits land and its max wind speed upon landfall
# mouse hover over point to show hurricane's identifier and year
htrack = alt.Chart(df, width=900, height=800).mark_circle().encode(
    latitude='latitude:Q',
    longitude='longitude:Q',
    color=alt.Color('max_wind', scale=alt.Scale(scheme='reds')),
    tooltip=['identifier', 'year(datetime)']
).transform_filter((alt.datum.record_id == 'L') & (alt.datum.status == 'HU') & (alt.datum.identifier != 'AL012017'))
states2 + htrack
```

```
Out[39]:
```



## Filtered dataset with only Hurricanes and their Landfall Data Points

```
In [40]: hurricanes = df[df['status']=='HU']
```

```
In [41]: hurricanes_landfall = hurricanes[hurricanes['record_id']=='L']
```

```
In [42]: hurricanes_landfall
```

	identifier	name	num_pts	record_id	status	latitude	longitude	max_wind	min_pressure	datetime
4	AL011851	UNNAMED	14	L	HU	28.2	-96.8	80	-999	1851-06-25T21:00:00
48	AL041851	UNNAMED	49	L	HU	30.1	-85.7	100	-999	1851-08-23T21:00:00
127	AL011852	UNNAMED	45	L	HU	30.2	-88.6	100	961	1852-08-26T06:00:00
163	AL031852	UNNAMED	20	L	HU	28.0	-82.8	70	-999	1852-09-12T00:00:00
223	AL051852	UNNAMED	25	L	HU	29.9	-84.4	90	-999	1852-10-09T21:00:00
...	...	...	...	...	...	...	...	...	...	...
50062	AL132017	KATIA	18	L	HU	20.6	-97.1	65	990	2017-09-09T03:00:00
50139	AL152017	MARIA	68	L	HU	15.4	-61.3	145	922	2017-09-19T01:15:00
50146	AL152017	MARIA	68	L	HU	18.0	-65.9	135	920	2017-09-20T10:15:00
50214	AL162017	NATE	32	L	HU	29.1	-89.2	75	983	2017-10-08T00:00:00
50215	AL162017	NATE	32	L	HU	30.4	-89.0	65	984	2017-10-08T05:20:00

498 rows × 10 columns