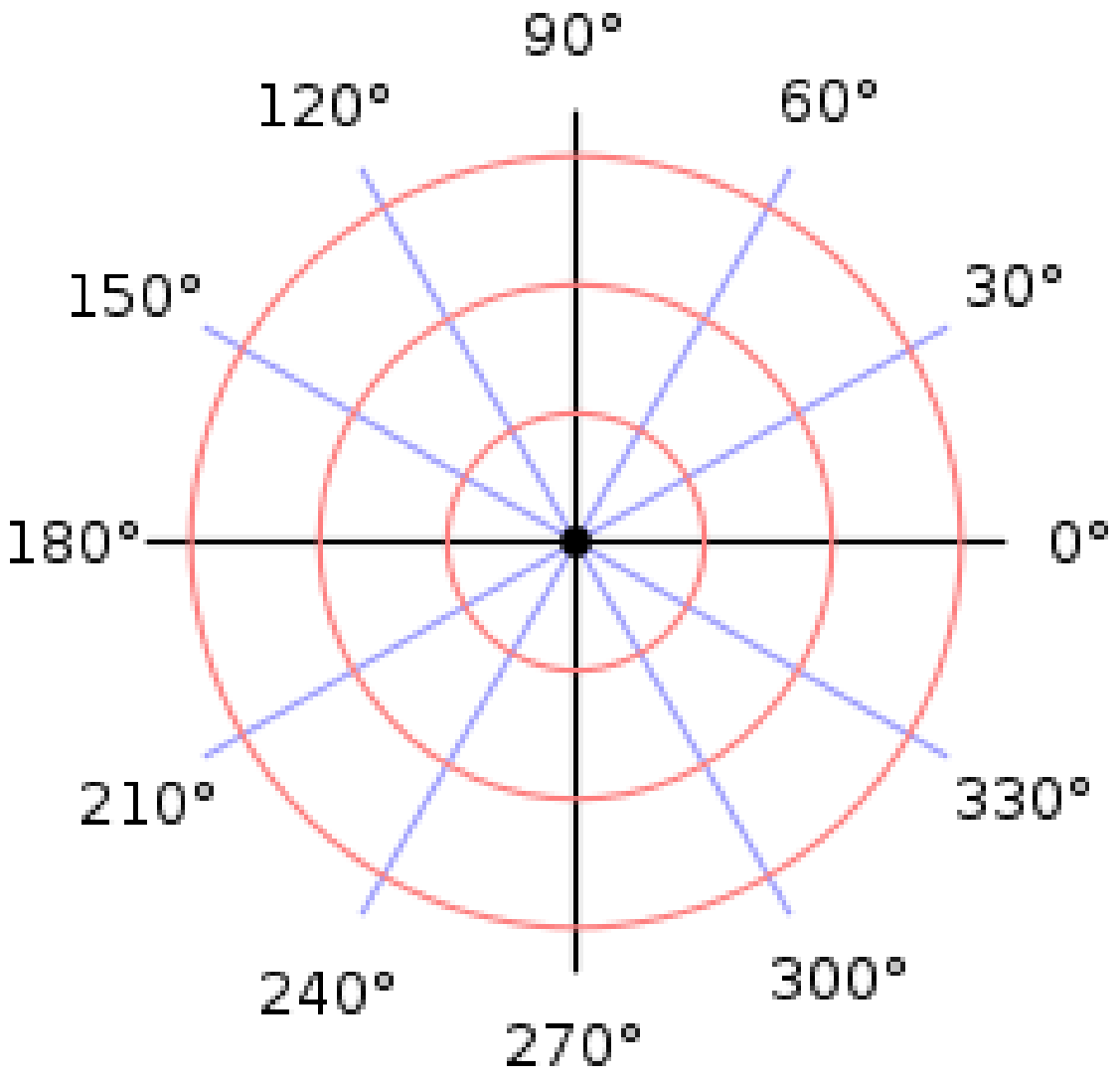
Directional Statistics

Directional statistics is a branch of statistics that deals with data measured in angles, directions, or orientations, focusing on the analysis and interpretation of cyclical or circular data.

Basic Concepts



Directional and circular data used interchangeably in analysis.

Circular data is data that is measured on a circle in degrees or radians. It is fundamentally different from linear data due to its periodic nature (0° = 360°).

Directional/circular distributions: No true zero, high/low values arbitrary.

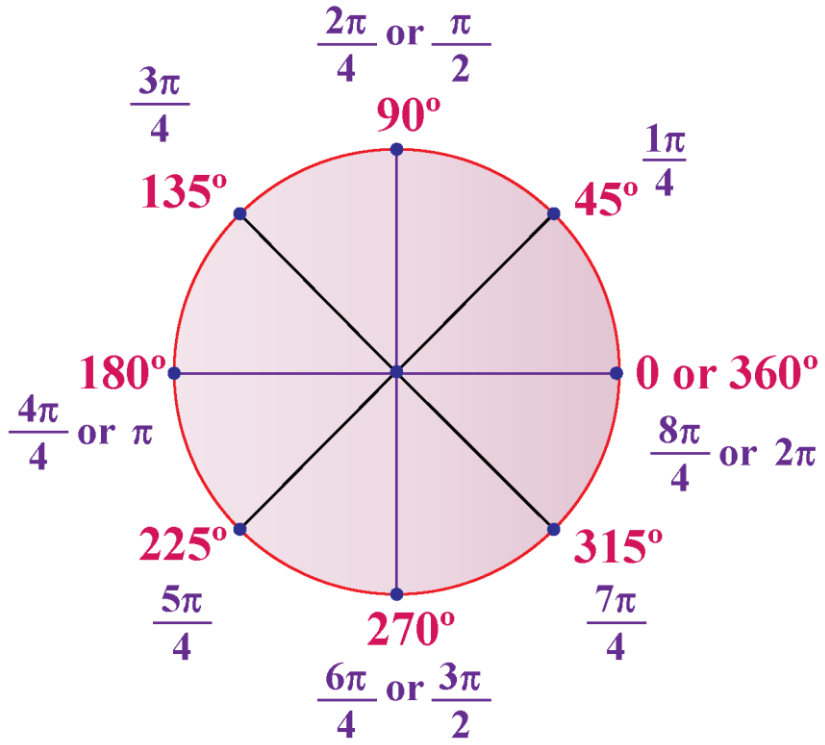
Analysis based on the polar coordinate system.

Note: Directional analysis also possible on Cartesian Coordinate System (requires angle calculation).

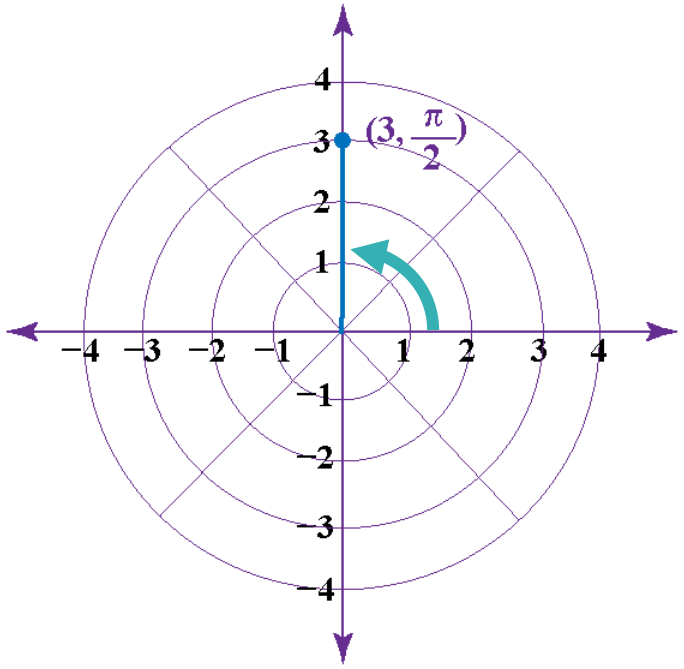
Polar Coordinate System:

Three key elements: pole, angle, radical distance

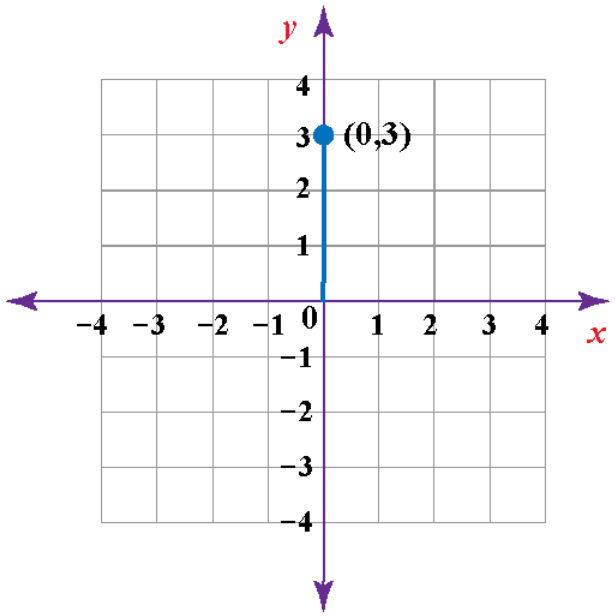
- The pole: the reference point (i.e. the point of origin in the Cartesian Coordinate System)
- Angle : indicating the **directional** distance of a point from the polar axis
    - *Angle can be in Degree or in Radian*
    - *Degree to Radian : ° × π/180*
- Radical distance/radius: distance from the pole



A comparison between expression under the Polar Coordinate System and the Cartesian Coordinate System:

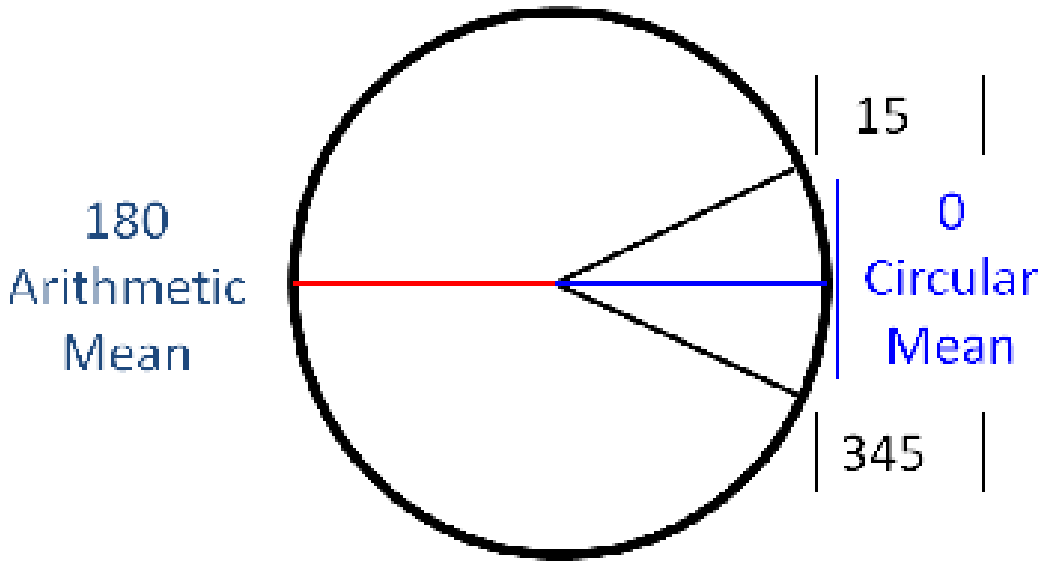


Polar Grid



Coordinate Grid

What are the basics in circular statistics?



- $\boldsymbol{Circular\ Mean\ (unit) = \arctan\left(\frac{\sum_{i=1}^{n} sin\ \theta}{\sum_{i=1}^{n} cos\ \theta}\right)}$

$$Circular\ Mean\ (unit) = \arctan\frac{\sin 15° + \sin 345°}{\cos 15° + \cos 345°}$$

- $$\begin{cases} if: sin > 0,\ cos > 0, Circular\ Mean * \frac{PI}{180} \\ if: sin > 0, cos < 0, 360, Circular\ Mean * \frac{PI}{180} + 180 \\ if: sin < 0, cos > 0, 360, Circular\ Mean * \frac{PI}{180} + 360 \\ if: sin < 0, cos > 0, 360, Circular\ Mean * \frac{PI}{180} + 180 \end{cases}$$

Mean Resultant Length

- Mean Resultant Length (MRL, $\overline{R}$) =

$$\sqrt{(\frac{1}{n}\sum_{i=1}^{n}\sin\theta)^2+(\frac{1}{n}\sum_{i=1}^{n}\cos\theta)^2}$$

- The **larger** the value, the **more concentrated** is the data.
- Between 0 and 1.

Circular Variance

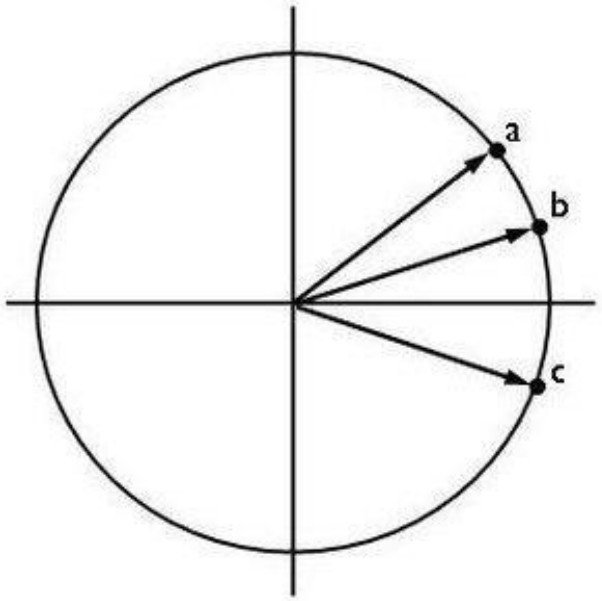
- Opposite to MRL: measures the degree of spread of the data

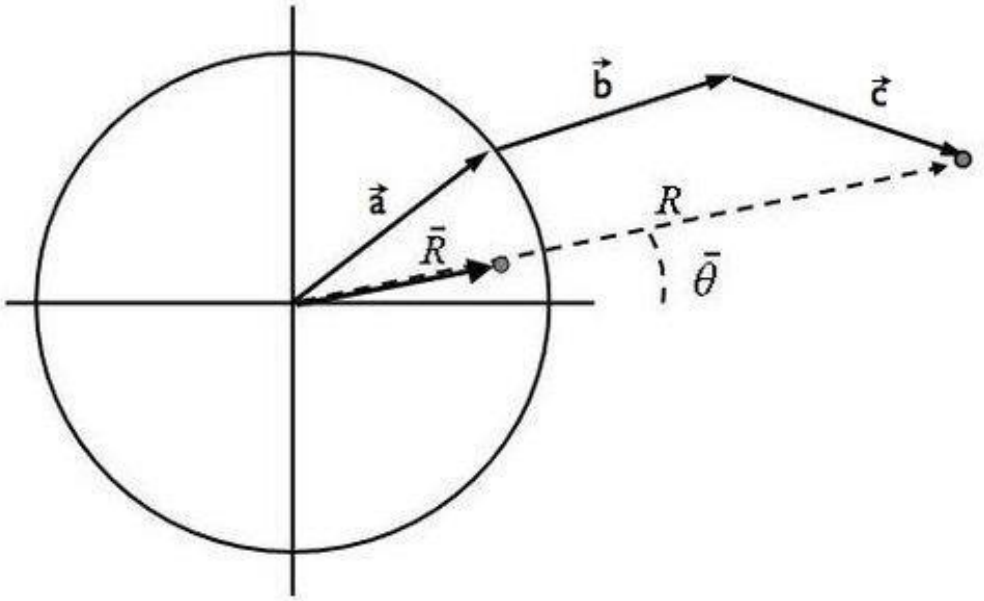
$$\textbf{Circular Variance} = 1 - \overline{R}$$

- Range between **0** and **1**
- The **larger** the value, the **more spread** are the data points.

Real Life:
wind speed +wind direction

- $Directional\ Mean\ (unit) = \arctan \frac{\sum_{i=1}^{n} \text{windspeed}_i * \sin(\theta_i)}{\sum_{i=1}^{n} \text{windspeed}_i * \cos(\theta_i)}$





$Linear\ Mean$

- The linear mean is a **non-directional** calculation of the average of a set of numbers
- $Linear\ Mean\ (unit) = \frac{\sum_{i=1}^{n} \theta}{n}$

Application

- We are going to calculate and validate the monthly prevailing wind for Calgary in 2022

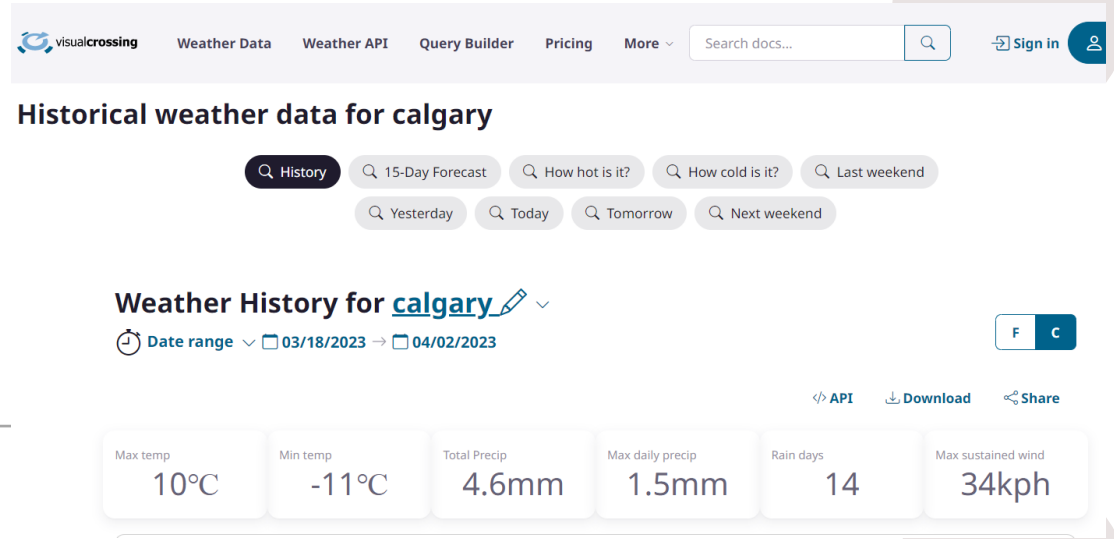
➔using circular mean

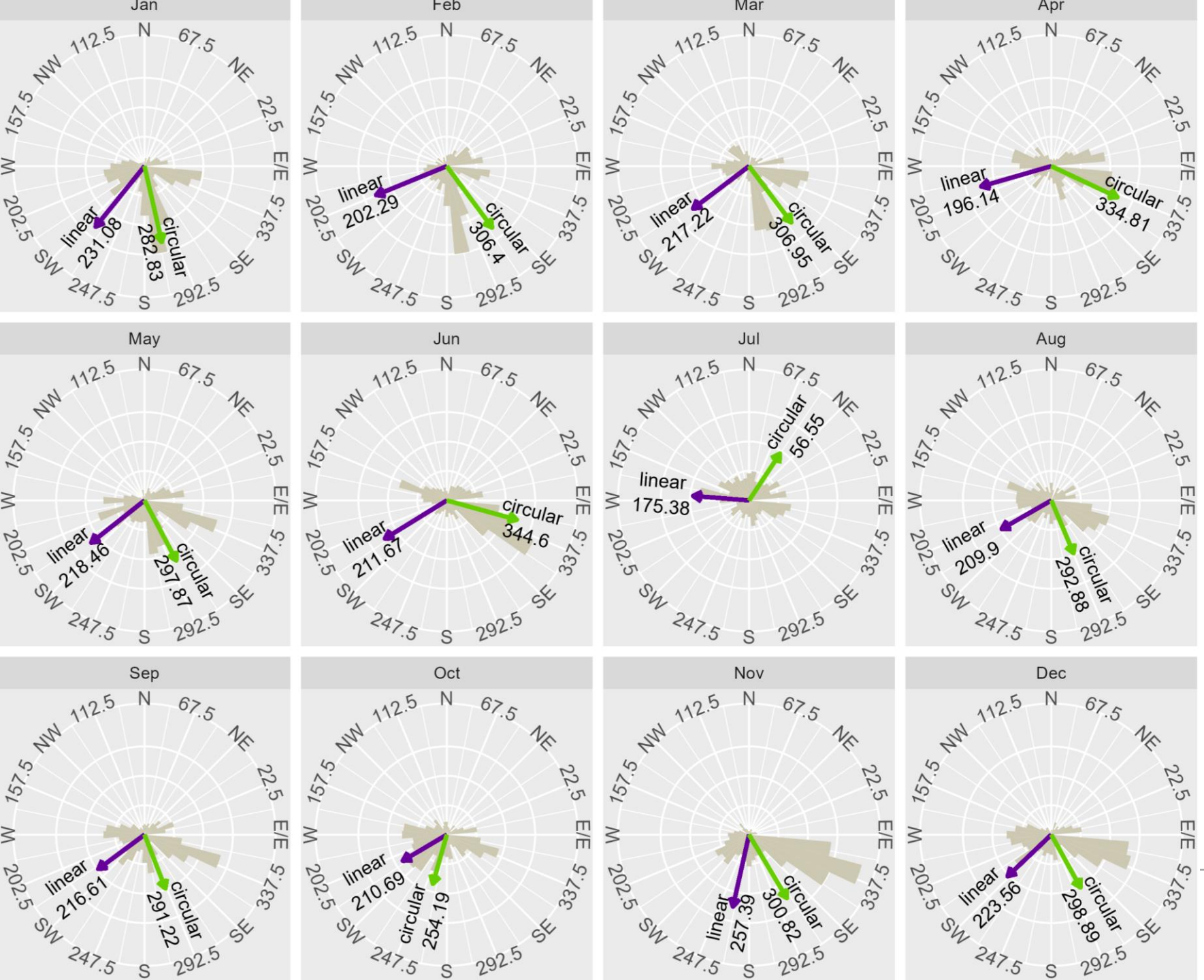
➔circular variance

➔real-world consideration

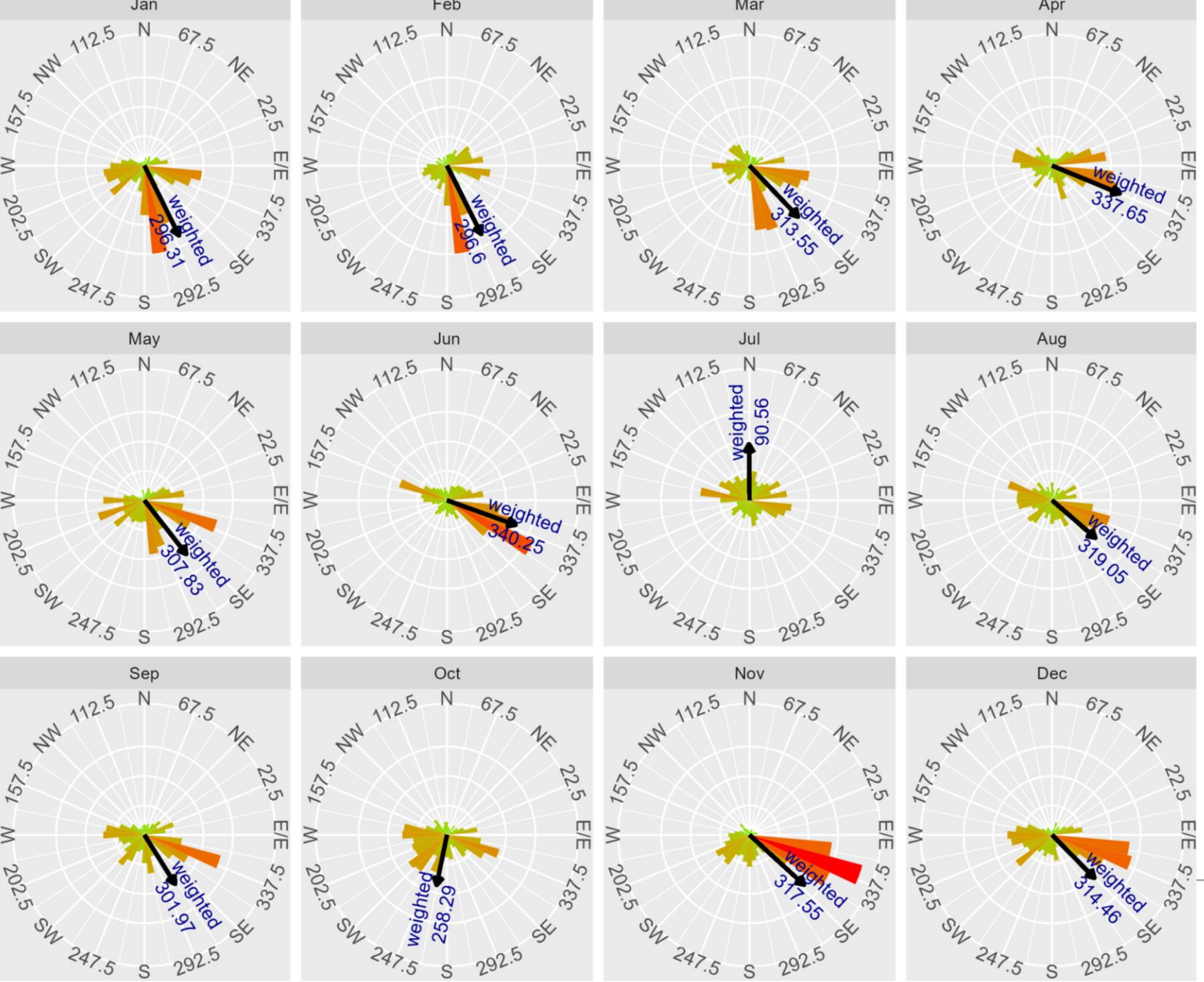
Application – Wind Direction

- Date: Jan 1, 2022 to Dec 31, 2022 (Hourly Weather Data)
- Location: Calgary
- Source: Weather History for Calgary from **Visual Crossing Weather Data**
  - Variables used:
  - Wind Direction
  - Wind Speed
  - Date





The Difference Between and Liner and Circular Value



Wind Direction by Circular Mean

Wind Direction Recorded hourly

Direction 0 25 50 75

Uniformity Test - Rayleigh's Test

- Null Hypothesis:
- H_0 : Wind directions are randomly/uniformly distributed.

$$Z = n\,\overline{R}^2$$

- Where n is the total number of observation.

- Reject the null hypothesis when $Z > Z_{critical}$, p-value < 0.005

Critical *z* Values for the Rayleigh's Test *Taken from Zar, 1981 Table B.32*

n	α: 0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
6	0.734	1.639	2.274	2.865	3.576	4.058	4.491	4.985	5.297
7	0.727	1.634	2.278	2.885	3.627	4.143	4.617	5.181	5.556
8	0.723	1.631	2.281	2.899	3.665	4.205	4.710	5.322	5.743
9	0.719	1.628	2.283	2.910	3.694	4.252	4.780	5.430	5.885
10	0.717	1.626	2.285	2.919	3.716	4.289	4.835	5.514	5.996
11	0.715	1.625	2.287	2.926	3.735	4.319	4.879	5.582	6.085
12	0.713	1.623	2.288	2.932	3.750	4.344	4.916	5.638	6.158
13	0.711	1.622	2.289	2.937	3.763	4.365	4.947	5.685	6.219
14	0.710	1.621	2.290	2.941	3.774	4.383	4.973	5.725	6.271
15	0.709	1.620	2.291	2.945	3.784	4.398	4.996	5.759	6.316
16	0.708	1.620	2.292	2.948	3.792	4.412	5.015	5.789	6.354
17	0.707	1.619	2.292	2.951	3.799	4.423	5.033	5.815	6.388
18	0.706	1.619	2.293	2.954	3.806	4.434	5.048	5.838	6.418
19	0.705	1.618	2.293	2.956	3.811	4.443	5.061	5.858	6.445
20	0.705	1.618	2.294	2.958	3.816	4.451	5.074	5.877	6.469
21	0.704	1.617	2.294	2.960	3.821	4.459	5.085	5.893	6.491
22	0.704	1.617	2.295	2.961	3.825	4.466	5.095	5.908	6.510
23	0.703	1.616	2.295	2.963	3.829	4.472	5.104	5.922	6.528
24	0.703	1.616	2.295	2.964	3.833	4.478	5.112	5.935	6.544
25	0.702	1.616	2.296	2.966	3.836	4.483	5.120	5.946	6.559
26	0.702	1.616	2.296	2.967	3.839	4.488	5.127	5.957	6.573
27	0.702	1.615	2.296	2.968	3.842	4.492	5.133	5.966	6.586
28	0.701	1.615	2.296	2.969	3.844	4.496	5.139	5.975	6.598
29	0.701	1.615	2.297	2.970	3.847	4.500	5.145	5.984	6.609
30	0.701	1.615	2.297	2.971	3.849	4.504	5.150	5.992	6.619
32	0.700	1.614	2.297	2.972	3.853	4.510	5.159	6.006	6.637
34	0.700	1.614	2.297	2.974	3.856	4.516	5.168	6.018	6.654
36	0.700	1.614	2.298	2.975	3.859	4.521	5.175	6.030	6.668
38	0.699	1.614	2.298	2.976	3.862	4.525	5.182	6.039	6.681
40	0.699	1.613	2.298	2.977	3.865	4.529	5.188	6.048	6.692
42	0.699	1.613	2.298	2.978	3.867	4.533	5.193	6.056	6.703
44	0.698	1.613	2.299	2.979	3.869	4.536	5.198	6.064	6.712
46	0.698	1.613	2.299	2.979	3.871	4.539	5.202	6.070	6.721
48	0.698	1.613	2.299	2.980	3.873	4.542	5.206	6.076	6.729
50	0.698	1.613	2.299	2.981	3.874	4.545	5.210	6.082	6.736
55	0.697	1.612	2.299	2.982	3.878	4.550	5.218	6.094	6.752
60	0.697	1.612	2.300	2.983	3.881	4.555	5.225	6.104	6.765
65	0.697	1.612	2.300	2.984	3.883	4.559	5.231	6.113	6.776
70	0.696	1.612	2.300	2.985	3.885	4.562	5.235	6.120	6.786
75	0.696	1.612	2.300	2.986	3.887	4.565	5.240	6.127	6.794
80	0.696	1.611	2.300	2.986	3.889	4.567	5.243	6.132	6.801
90	0.696	1.611	2.301	2.987	3.891	4.572	5.249	6.141	6.813
100	0.695	1.611	2.301	2.988	3.893	4.575	5.254	6.149	6.822
120	0.695	1.611	2.301	2.990	3.896	4.580	5.262	6.160	6.837
140	0.695	1.611	2.301	2.990	3.899	4.584	5.267	6.168	6.847
160	0.695	1.610	2.301	2.991	3.900	4.586	5.271	6.174	6.855
180	0.694	1.610	2.302	2.992	3.902	4.588	5.274	6.178	6.861
200	0.694	1.610	2.302	2.992	3.903	4.590	5.276	6.182	6.865
300	0.694	1.610	2.302	2.993	3.906	4.595	5.284	6.193	6.879
500	0.694	1.610	2.302	2.994	3.908	4.599	5.290	6.201	6.891
∞	0.6931	1.6094	2.3026	2.9957	3.9120	4.6052	5.2983	6.2146	6.9078

2.9957

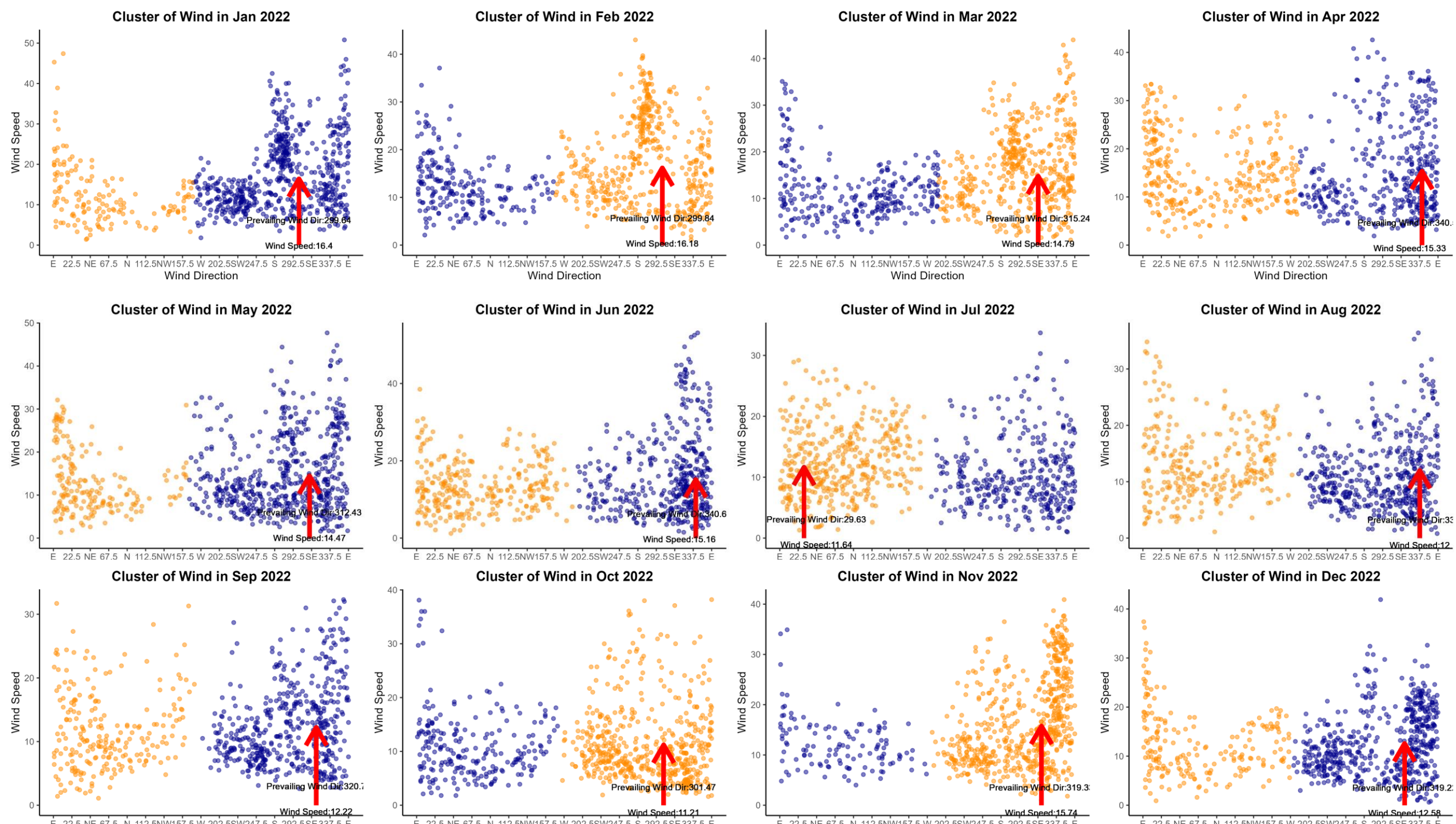
$Z_{critical}$

```
  mutate(month = month(datetime))
ifelse(unique(wind$MRL)>2.9957,"PASS","REJECT")
```

```
[1] "PASS" "PASS" "PASS" "PASS" "PASS" "PASS" "PASS" "PASS" "PASS" "PASS" "PASS" "PASS"
```

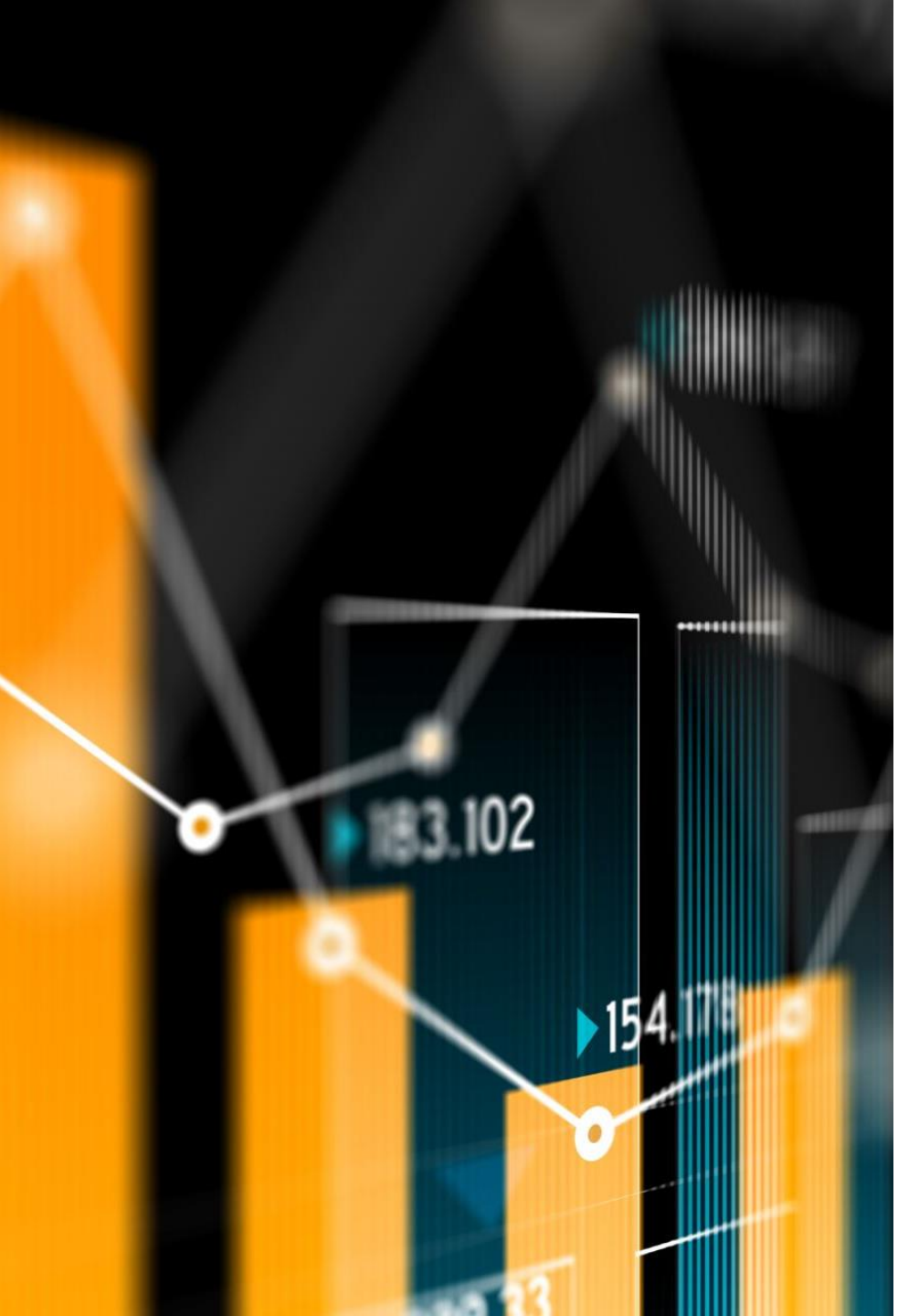
Rayleigh’s Test Results

P-Values < 0.05


Cluster of Wind in Jan 2022
Prevailing Wind Dir:299.64
Wind Speed:16.4
Cluster of Wind in Feb 2022
Prevailing Wind Dir:299.84
Wind Speed:16.18
Cluster of Wind in Mar 2022
Prevailing Wind Dir:315.24
Wind Speed:14.79
Cluster of Wind in Apr 2022
Prevailing Wind Dir:340.
Wind Speed:15.33
Cluster of Wind in May 2022
Prevailing Wind Dir:312.43
Wind Speed:14.47
Cluster of Wind in Jun 2022
Prevailing Wind Dir:340.6
Wind Speed:15.16
Cluster of Wind in Jul 2022
Prevailing Wind Dir:29.63
Wind Speed:11.64
Cluster of Wind in Aug 2022
Prevailing Wind Dir:33
Wind Speed:12
Cluster of Wind in Sep 2022
Prevailing Wind Dir:320.7
Wind Speed:12.22
Cluster of Wind in Oct 2022
Prevailing Wind Dir:301.47
Wind Speed:11.21
Cluster of Wind in Nov 2022
Prevailing Wind Dir:319.3
Wind Speed:15.74
Cluster of Wind in Dec 2022
Prevailing Wind Dir:319.2
Wind Speed:12.58
Wind Speed
Wind Direction
E 22.5 NE 67.5 N 112.5NW157.5 W 202.5SW247.5 S 292.5SE337.5 E

Demonstration Time

- Let’s have a try!
- Go to Github, get the dataset and get your hands dirty!



Conclusion

1. Directional statistics are vital for analyzing cyclical data in real-life applications, including higher-dimensional scenarios.
2. The versatility of directional statistics extends to various contexts beyond geometric objects, such as time and periodic measurements.

Solution for Error Run Directional

- 1. Install or update the OpenGL library: For macOS, you will need to install XQuartz. You can download and install it from the following website: https://www.xquartz.org/
- 2. Once the installation is complete, please restart your computer to ensure the changes take effect. Make sure you are using the latest version of R and the relevant packages. You can update all installed packages using the following code: update.packages(ask = FALSE)