



# **Uncovering the Impact of Climate on Indego Bike Rentals**

**Shannon He, Yuna Choi**

# Introduction



- **Indego** is a bicycle-sharing system with over **250 Stations** and **2000 Bikes** in Philadelphia
- Users could rent bikes and commute across the city

- **Aim**

To explore the relationship between Indego bike usage patterns and weather factors in Philadelphia



# PHILADELPHIA WEATHER Dataset



- From **Visual Crossing Corporation**
- Daily weather data **in 2023**
- A total of **365 observations** with **33 variables**

name, datetime, tempmax, tempmin, **temp**, feelslikemax, feelslikemin, **feelslike**, dew, humidity, **precipitation**, precipprob, precipcover, preciptype, snow, snowdepth, windgust, **windspeed**, winddir, sealevelpressure, cloudcover, **visibility**, solarradiation, **solarenergy**, uvindex, **severerisk**, sunrise, sunset, moonphase, conditions, description, icon, stations ,count, duration, distance



# Indego Trip Dataset

- From **Indego website**
- Trip data in **2023**
- A total of **272,539 trips** with **15 variables**
- Convert to dataset with 365 observations
- Combine weather and indego dataset

Trip id	distance	date
trip1	2	2023/01/01
trip2	1	2023/01/01
trip3	1	2023/01/02
trip4	4	2023/01/02
trip5	1	2023/01/02
trip6	2	2023/01/02
trip7	3	2023/01/03

2

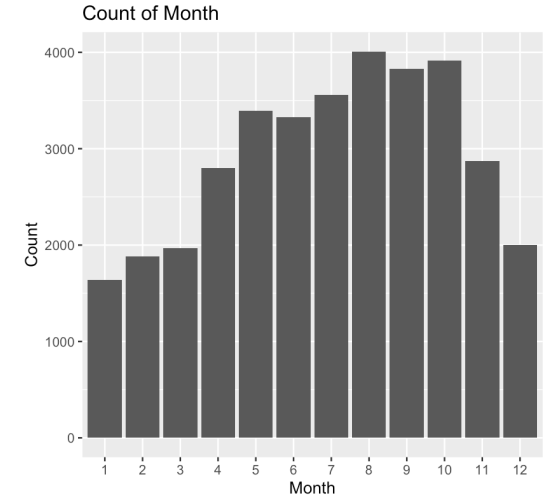
4

1

date	count	Total distance
2023/01/01	2	2+1 = 3
2023/01/02	4	1+4+1+2 = 8
2023/01/03	1	3

# INDEGO DATASET EDA

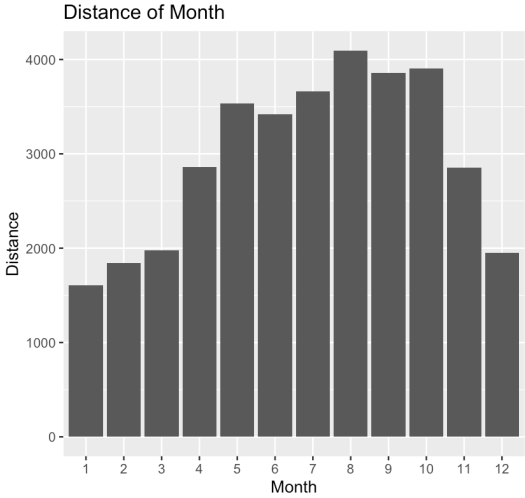
## COUNT



```
summary(raw$count)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
707	2073	3035	2938	3742	5160

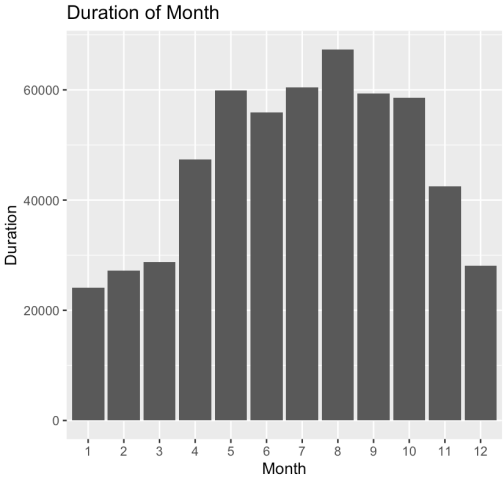
## DISTANCE



```
summary(raw$distance)
```

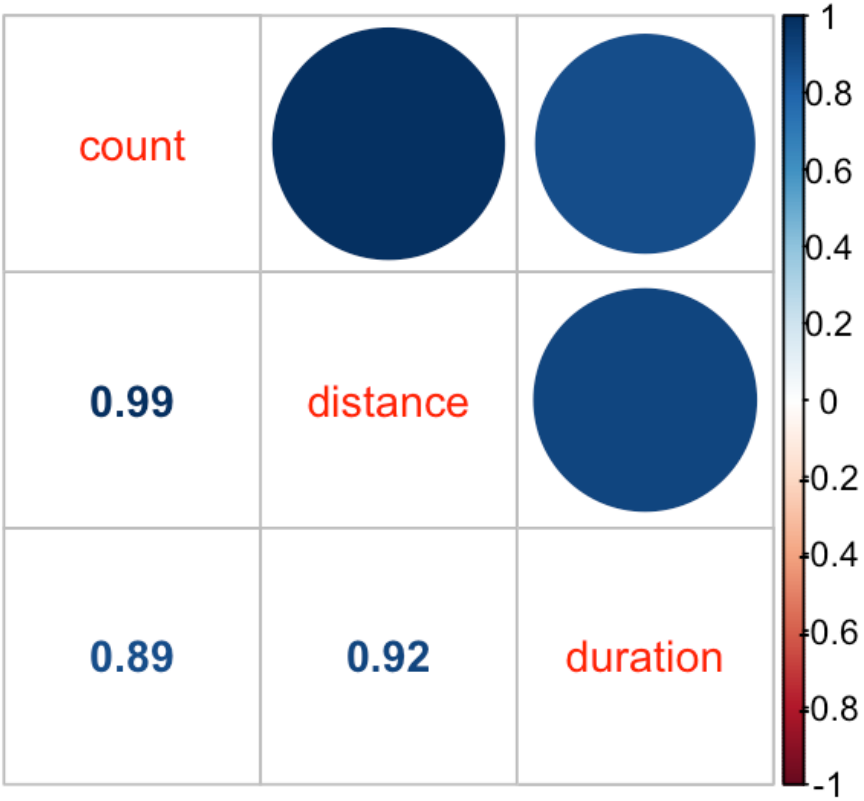
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
670.6	2053.1	3094.9	2970.3	3845.3	5356.2

## DURATION



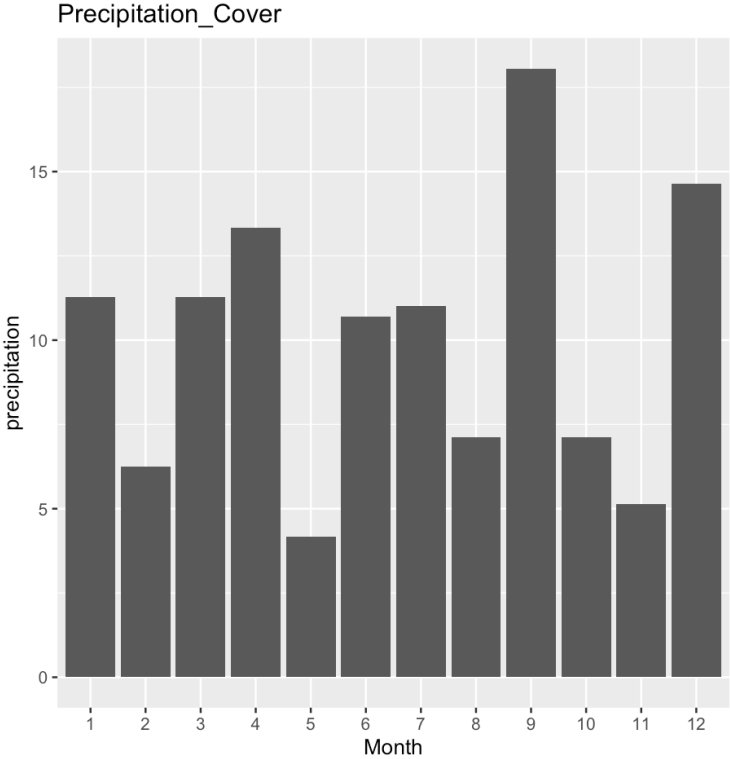
```
summary(raw$duration)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10908	29099	47659	46731	62880	105309



Chose 'count' as our response variable

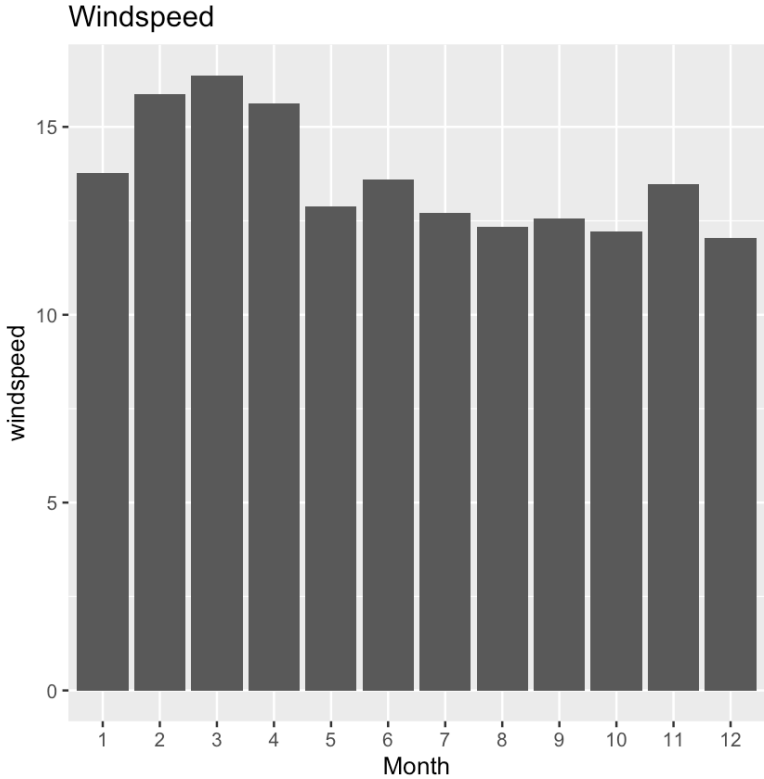
# PHILADELPHIA WEATHER EDA



**Precipitation\_cover**

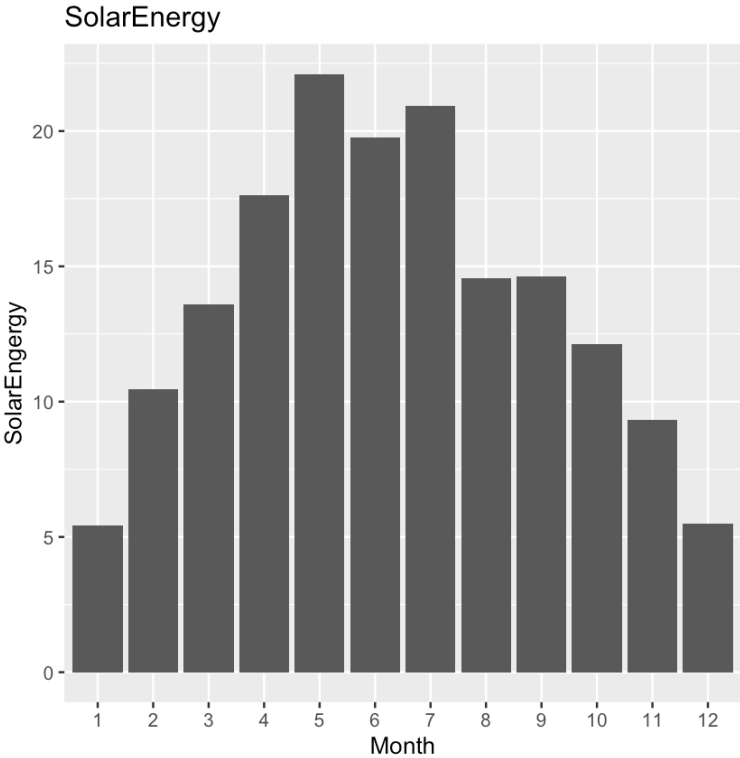
the proportion of hours where there was non-zero precipitation

Max- September      Min-May



**Windspeed**

February to April had the highest values.



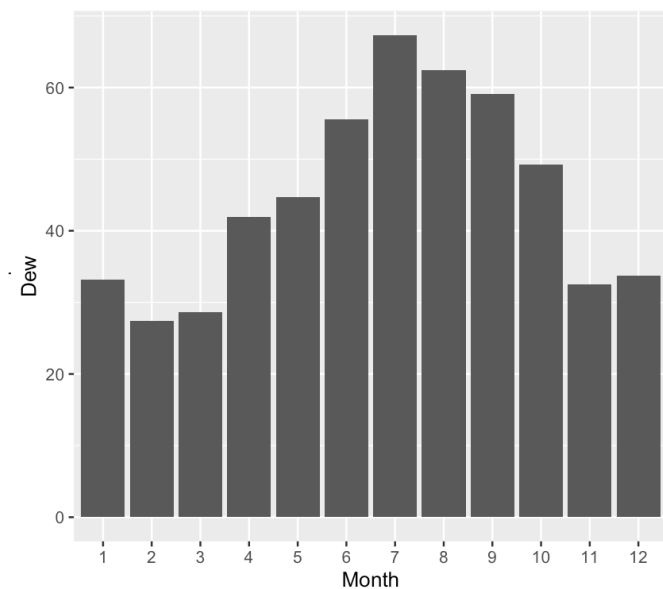
**Solar Energy**

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.423	10.172	14.066	13.836	18.165	22.113

Min – January      Max - May      6

# PHILADELPHIA WEATHER EDA

Dew of Month

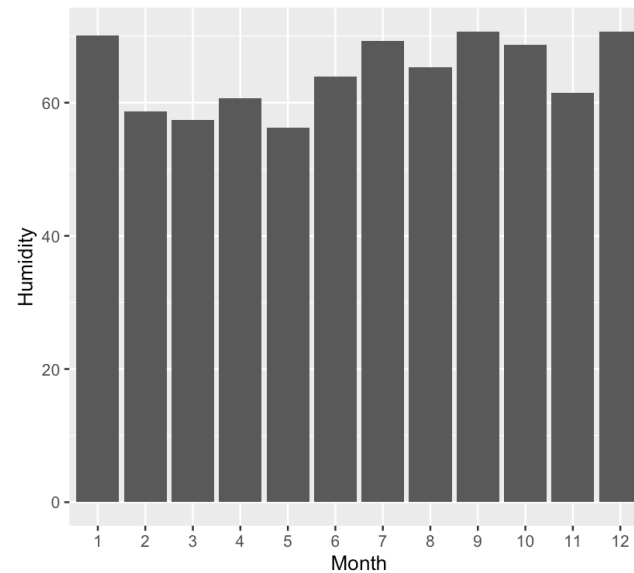


DEW

```
summary(raw$dew)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.20	32.40	44.70	44.77	57.90	74.70

Humidity of Month

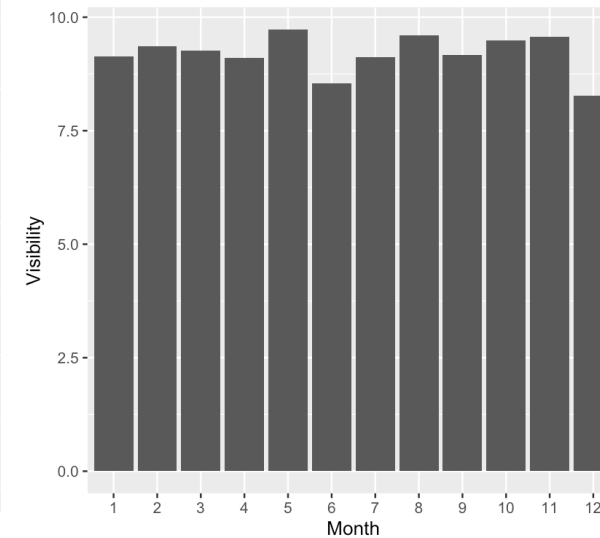


HUMIDITY

```
summary(raw$humidity)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
26.70	54.40	64.60	64.48	74.70	97.30

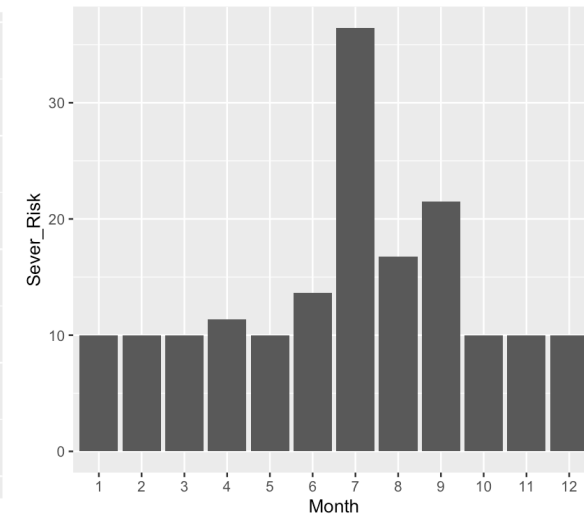
Visibility



Visibility

distance at which distant objects are visible

Sever\_Risk

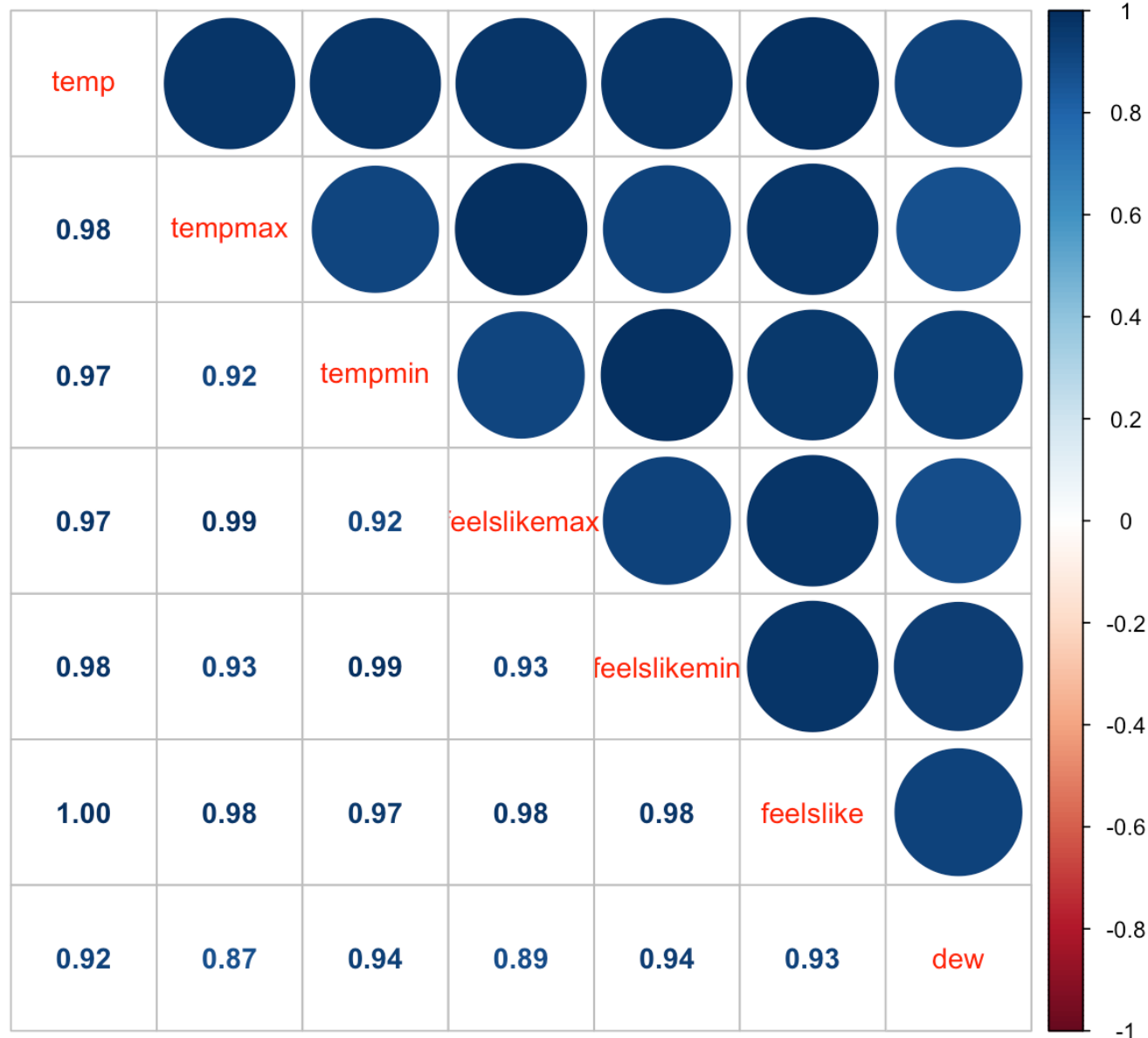


Severe risk

a value representing the risk of convective storms (e.g. thunderstorms, hail and tornadoes)

Max-July

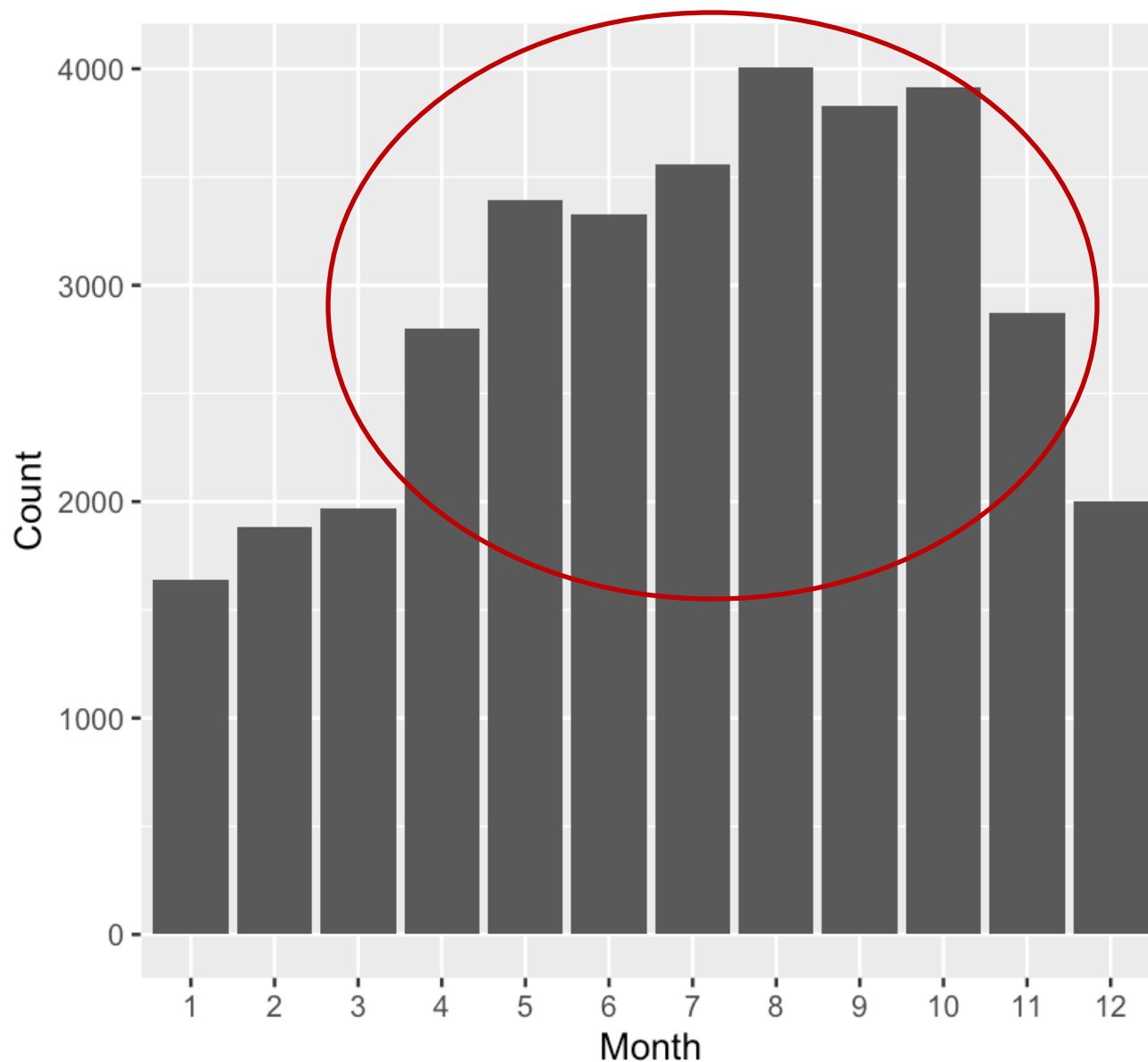
# Temperature Related Variables



- **Seven variables** related to temperature
  - Temperature (Max, Min and Average)
  - Feels like temperature (Max, Min and Average)
  - Dew point
- **High correlation** between variables
- Keep only one variable (**Feels like temperature**) for our variable selection candidates



## Add Season Variable



- There are **high bike rental count** from **May to October**
- Potential correlation between count and season
- Add season variable

# Multiple Regression Model

## Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-3.156	367.304	-0.009	0.993150	
feelslike	38.133	2.895	13.174	< 2e-16	***
precipcover	-15.268	2.077	-7.351	1.37e-12	***
windspeed	-14.652	6.289	-2.330	0.020378	*
visibility	72.767	24.324	2.992	0.002970	**
solarenergy	22.429	5.353	4.190	3.53e-05	***
seasonspring	-785.827	80.832	-9.722	< 2e-16	***
seasonsummer	-585.760	87.116	-6.724	7.11e-11	***
seasonwinter	-856.074	88.135	-9.713	< 2e-16	***
severerisk1low	770.866	130.164	5.922	7.52e-09	***
severerisk1moderate	521.738	147.116	3.546	0.000443	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 485.7 on 354 degrees of freedom

Multiple R-squared: 0.7914, Adjusted R-squared: 0.7855

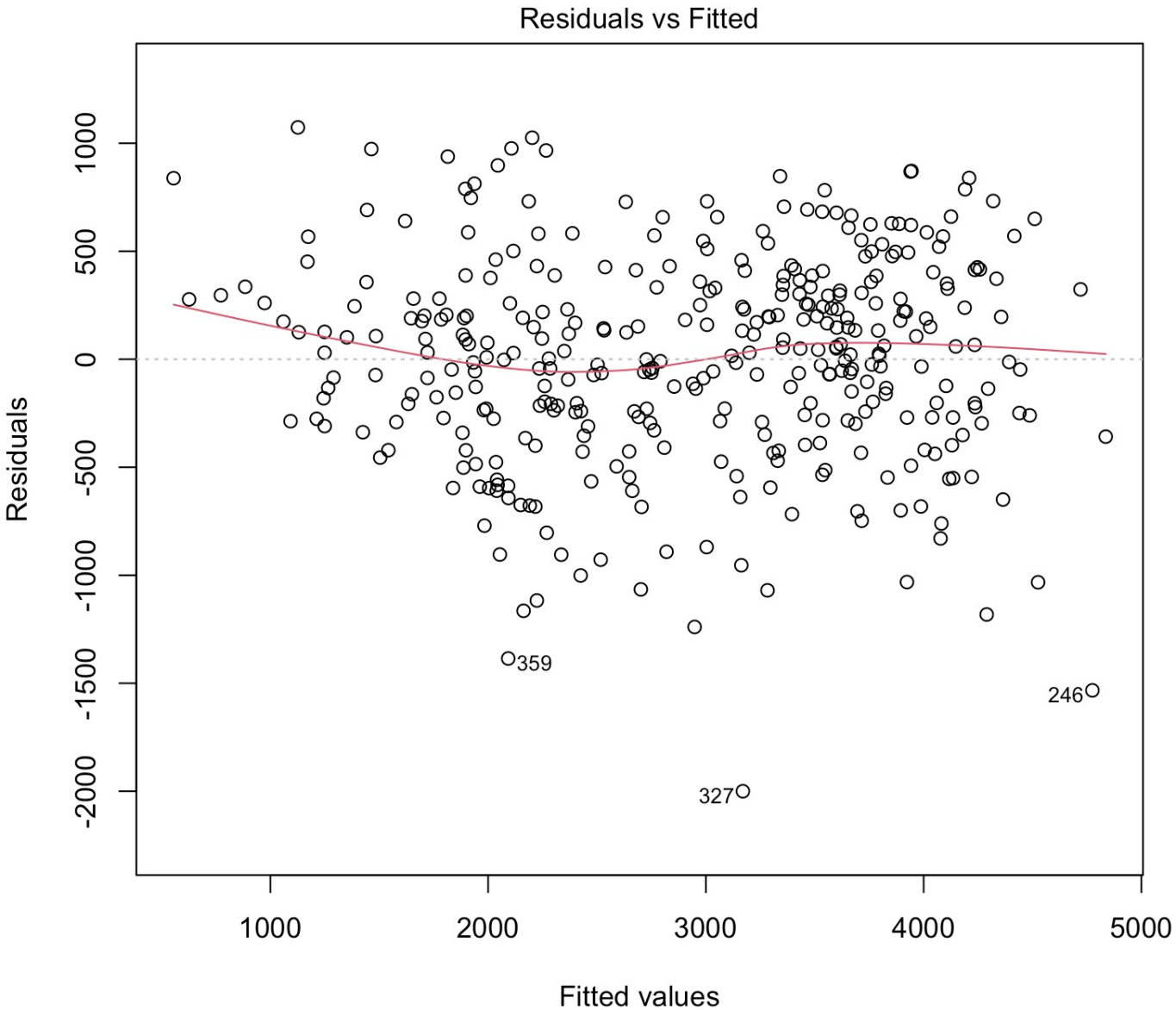
F-statistic: 134.3 on 10 and 354 DF, p-value: < 2.2e-16

- Stepwise regression
- Response: **Daily rental count**
- **7** independent variables
- R squared = 0.7914
- Explain **79.14%** of variability in data

## • High bike rental

- **Fall**>summer> spring>winter
- Low **severe risk**
- Higher **feels like temperature**
- Higher **visibility**
- Higher **solar energy**
- Lower **precipitation cover**
- Lower **wind speed**

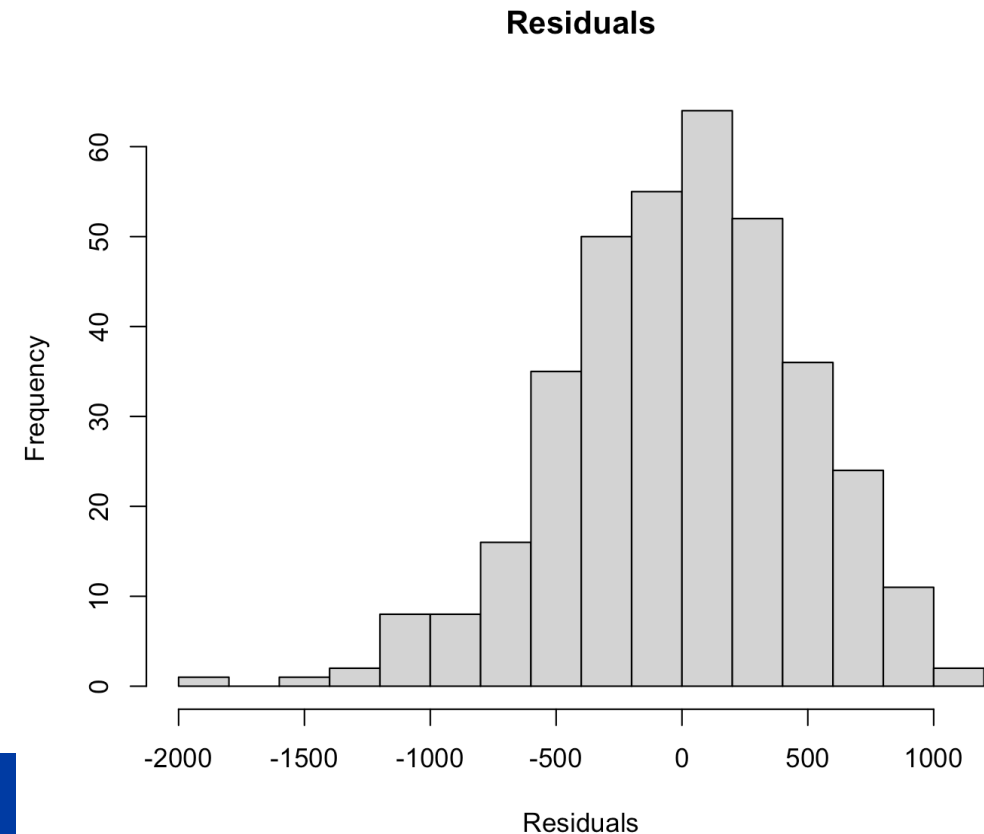
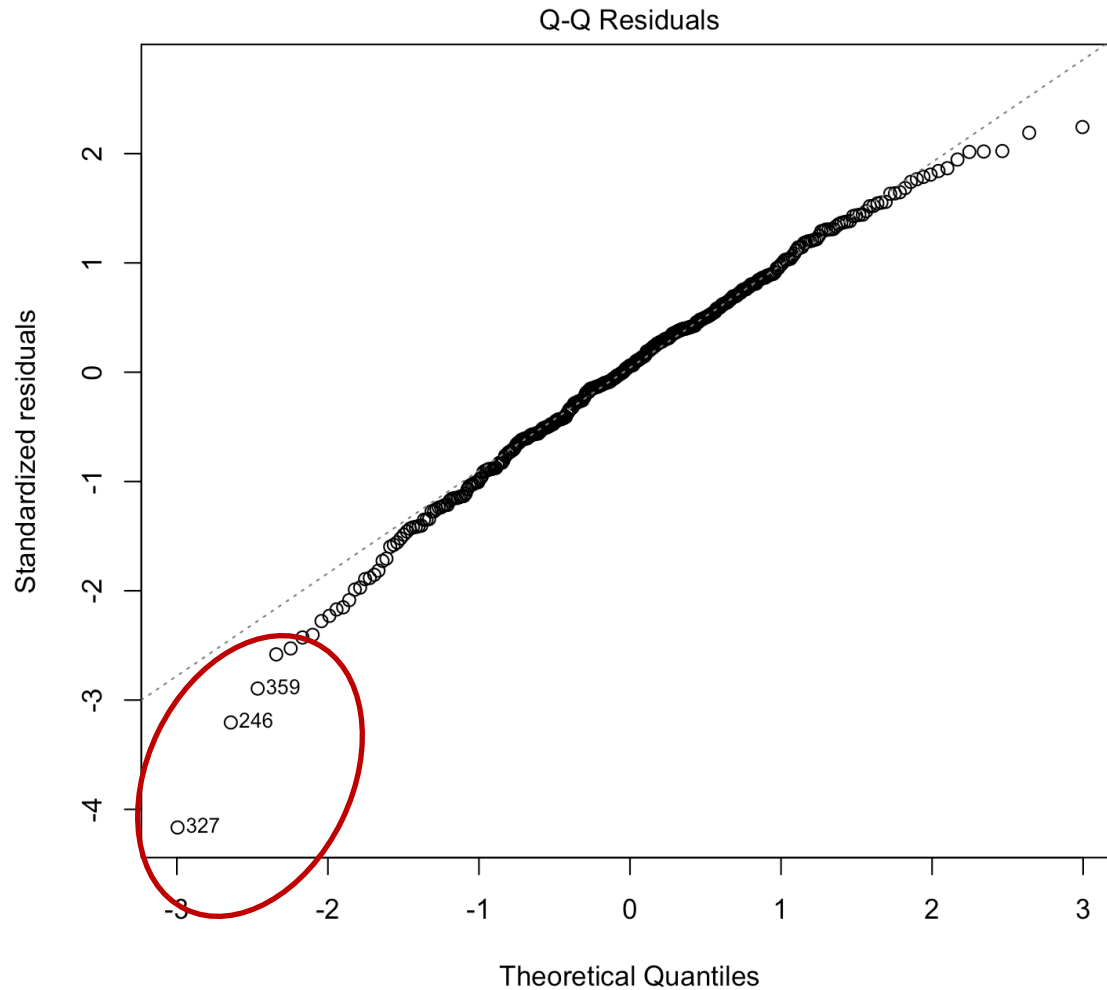
# Residuals vs Fitted Plot



- Residuals equally spread around the horizontal 0 line
- **Without a distinct pattern**
- The variances of residuals are **equal**

# The Normality Assumption

- Whether data is normally distributed
- **Both ends lie below** the reference line
- **Left skewness** in the residuals
- May be caused by **outliers**



# Conclusion

- Stepwise regression model to investigate **linear relationship** between daily bike rentals and weather-related variables
- Weather-related variables could explain **79.14%** of variability within the data
- There are more daily bike rentals in the **fall**
- **Warm weather conditions** were associated with higher bike rentals
- Include other variables may improve the model
  - **Weekday/Weekend variable**
  - Activities held in Philly