



...

Seattle Weather Analysis: Rain, Sun, or Snow?



Andy Chen



Understanding Weather Patterns ...

Weather forecasting is crucial for planning and managing daily activities, especially in a city like Seattle known for its variable climate. This project uses a **Generalized Logit Model for Nominal Response** to analyze how temperature, precipitation, and wind influence the likelihood of rain, sun, or snow.



... Dataset

This dataset, sourced from Seattle weather records between 2012 and 2015, contains 24 observations with variables including **date**, **precipitation**, **max** and **min temperatures**, **wind speed**, and **weather type** (**rain**, **sun**, **snow**), aiming to predict the weather type based on numerical factors.

	A	B	C	D	E	F
1	date	precipitation	temp_max	temp_min	wind	weather
2	2012/1/1	0	12.8	5	4.7	rain
3	2012/1/2	10.9	10.6	2.8	4.5	rain
4	2012/1/3	0.8	11.7	7.2	2.3	rain
5	2012/1/4	20.3	12.2	5.6	4.7	rain
6	2012/1/5	1.3	8.9	2.8	6.1	rain
7	2012/1/6	2.5	4.4	2.2	2.2	rain
8	2012/1/7	0	7.2	2.8	2.3	rain
9	2012/1/8	0	10	2.8	2	sun
10	2012/1/9	4.3	9.4	5	3.4	rain
11	2012/1/10	1	6.1	0.6	3.4	rain
12	2012/1/11	0	6.1	-1.1	5.1	sun
13	2012/1/12	0	6.1	-1.7	1.9	sun
14	2012/1/13	0	5	-2.8	1.3	sun
15	2012/1/14	4.1	4.4	0.6	5.3	snow
16	2012/1/15	5.3	1.1	-3.3	3.2	snow
17	2012/1/16	2.5	1.7	-2.8	5	snow



SAS Code

```
proc import datafile="seattle-weather.csv"
  out=weather_data
  dbms=csv
  replace;
  getnames=yes;
run;

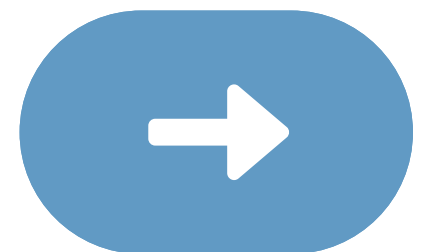
proc logistic data=weather_data descending;
  class weather (ref='sun') / param=ref; /* Set 'sun' as reference category */
  model weather = precipitation temp_max temp_min wind / link=glogit;
run;

data predict;
  input precipitation temp_max temp_min wind;
  cards;
100 75.2 52.3 4.5
;
run;

/* Combine the original and prediction datasets */
data weather_combined;
  set weather_data predict;
run;

/* Fit the Generalized Logit Model */
proc logistic data=weather_combined descending;
  class weather (ref='sun') / param=ref;
  model weather = precipitation temp_max temp_min wind / link=glogit;
  output out=weather_predictions p=pweather;
run;

proc print data=weather_predictions noobs;
  where precipitation=0.01 or precipitation=0.00 or precipitation=0.25;
  var precipitation temp_max temp_min wind _level_ pweather;
run;
```



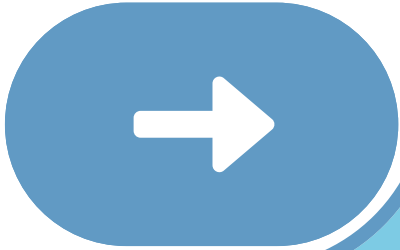
SAS Output

Analysis of Maximum Likelihood Estimates						
Parameter	weather	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	snow	1	-1.3267	0.9464	1.9651	0.1610
Intercept	rain	1	1.1021	0.3226	11.6696	0.0006
precipitation	snow	1	25.1774	11.9515	4.4379	0.0351
precipitation	rain	1	25.1058	11.9514	4.4128	0.0357
temp_max	snow	1	-0.3728	0.1200	9.6445	0.0019
temp_max	rain	1	-0.1488	0.0268	30.8170	<.0001
temp_min	snow	1	-0.2278	0.1249	3.3269	0.0682
temp_min	rain	1	0.1471	0.0366	16.1578	<.0001
wind	snow	1	0.1157	0.1460	0.6278	0.4282
wind	rain	1	-0.3146	0.0762	17.0328	<.0001

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2237.589	1004.975
SC	2248.162	1057.844
-2 Log L	2233.589	984.975

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1248.6136	8	<.0001
Score	567.7481	8	<.0001
Wald	112.4101	8	<.0001

precipitation	temp_max	temp_min	wind	_LEVEL_	pweather
0	12.8	5	4.7	sun	0.82347
0	12.8	5	4.7	snow	0.00102
0	12.8	5	4.7	rain	0.17551
0	7.2	2.8	2.3	sun	0.56577
0	7.2	2.8	2.3	snow	0.00707
0	7.2	2.8	2.3	rain	0.42716





R Code



```
library(nnet)

# Load the data
weather_data <- read.csv("seattle-weather.csv")

# Ensure the weather variable is a factor
weather_data$weather <- factor(weather_data$weather, levels = c("rain", "sun", "snow"))

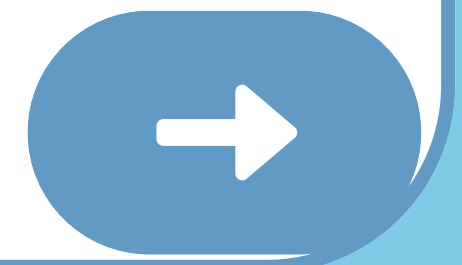
# Specify reference category for multinomial model
weather_rel <- relevel(weather_data$weather, ref = "sun") # Set "sun" as the reference category

# Fit the multinomial logistic regression model including wind as a predictor
fitted_model <- multinom(weather_rel ~ temp_max + temp_min + precipitation + wind, data = weather_data)

# Display the summary of the fitted model
summary(fitted_model)

ggplot(weather_data, aes(x = precipitation, y = temp_max)) +
  geom_point(aes(color = weather)) +
  geom_smooth(method = 'lm', se = FALSE, aes(color = "trend")) +
  theme_minimal() +
  labs(title = "Max Temperature vs. Precipitation by Weather Type", x = "Precipitation (mm)", y = "Max Temperature (°C)") +
  facet_wrap(~weather)

ggplot(weather_data, aes(x = weather, y = temp_max, fill = weather)) +
  geom_violin() +
  theme_minimal() +
  labs(title = "Max Temperature Distribution by Weather Type", x = "Weather Type", y = "Max Temperature (°C)") +
  theme(legend.position = "none")
```

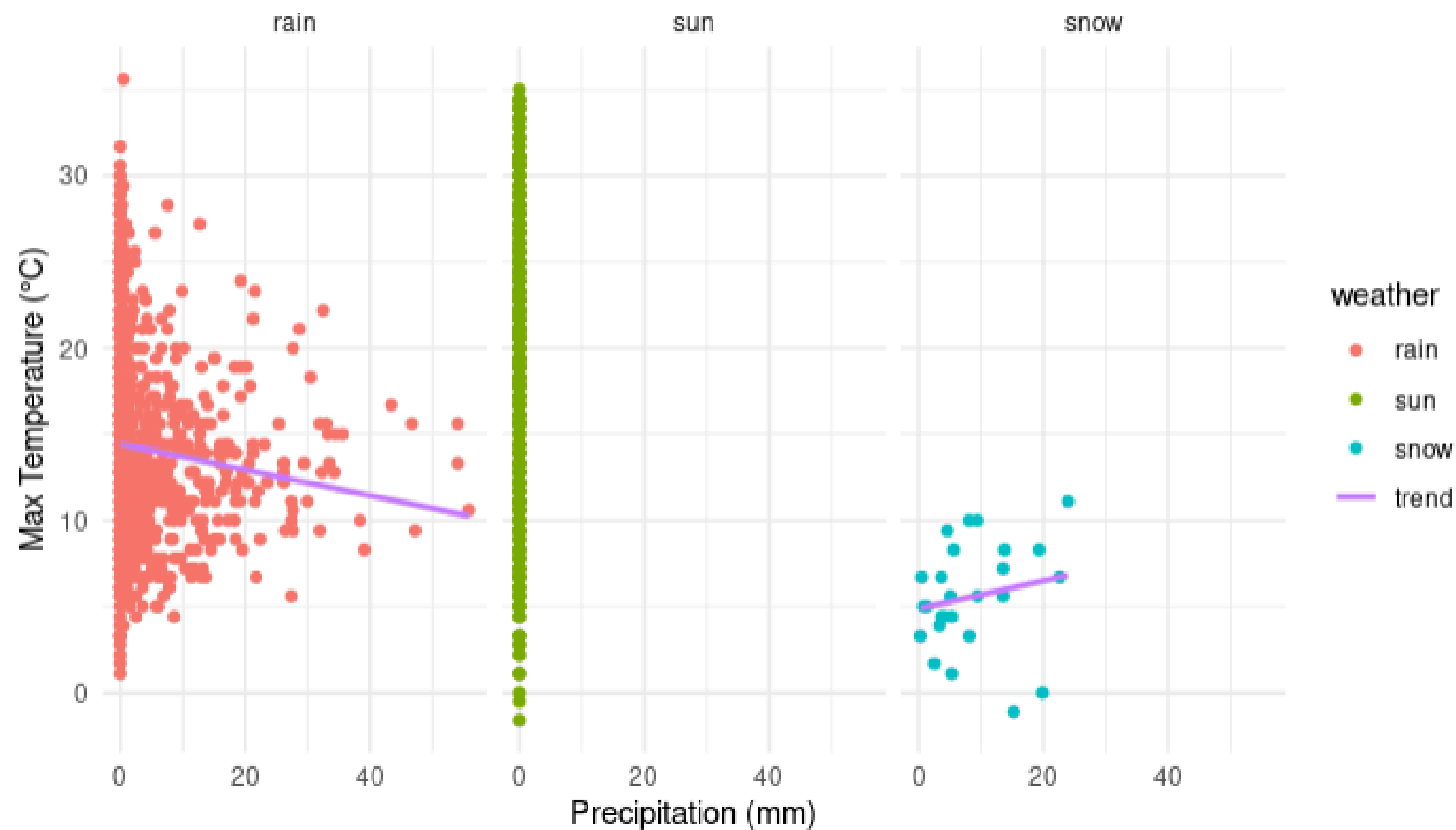


R Output

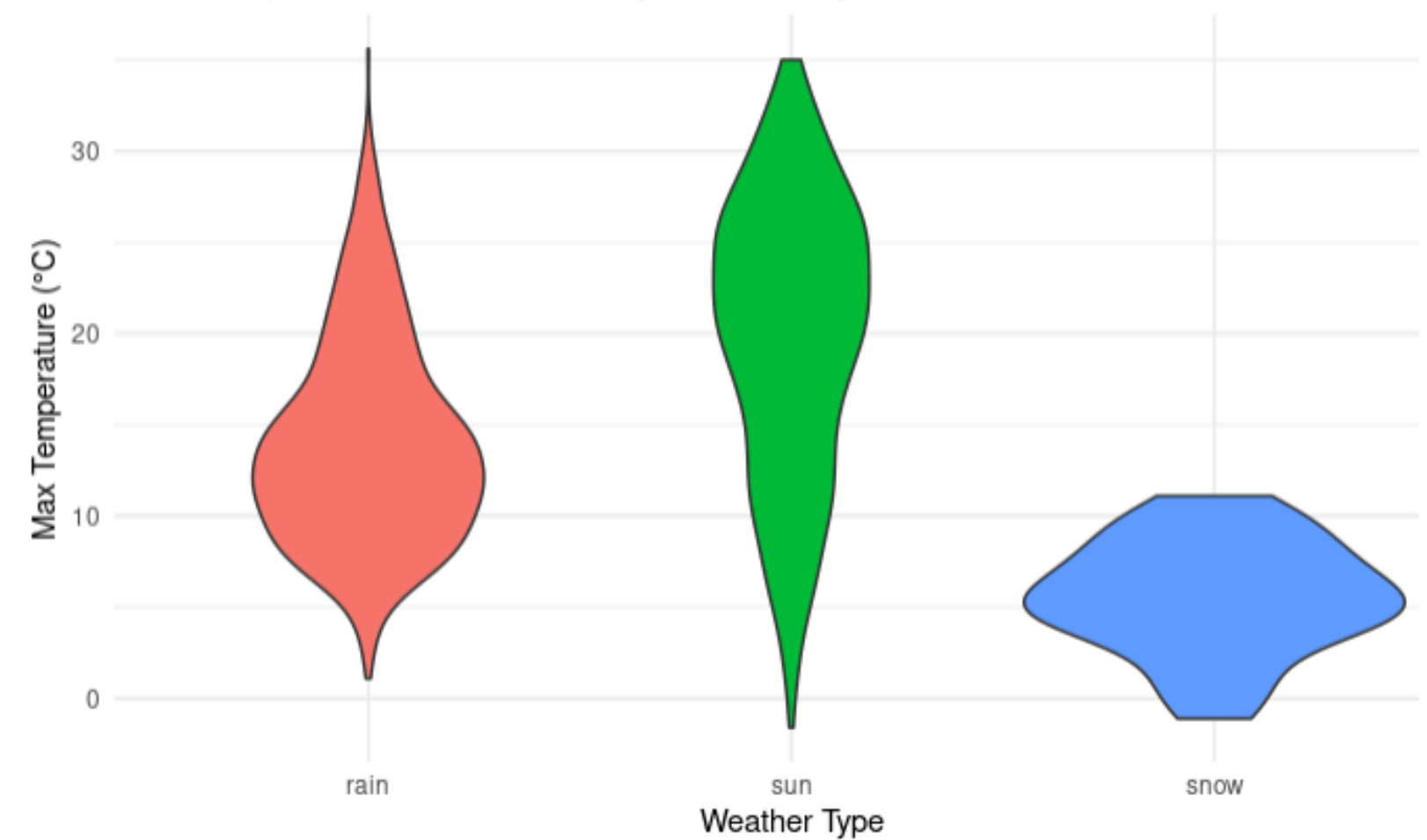
```
Coefficients:
      (Intercept)  temp_max  temp_min precipitation      wind
rain      1.102639 -0.1488128  0.1471095      33.07871 -0.3148382
snow     -1.326403 -0.3725706 -0.2276596      33.15028  0.1154182
```

```
Std. Errors:
      (Intercept)  temp_max  temp_min precipitation      wind
rain    0.3226910  0.02681287  0.0366069    0.01165899  0.07626059
snow    0.9461501  0.11997808  0.1248292    0.01165807  0.14594974
```

Max Temperature vs. Precipitation by Weather Type



Max Temperature Distribution by Weather Type





Fitted Model

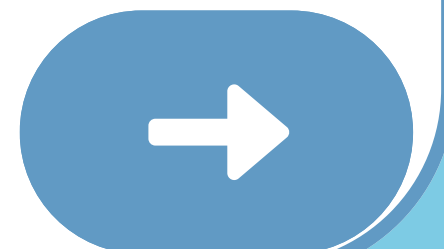


For rain versus sun

$$\lambda_{\text{rain}} = \exp(1.102639 + (-0.1488128(\text{temp_max})) + (0.1471095(\text{temp_min})) + (33.07871(\text{precipitation})) + (-0.3148382(\text{wind})))$$

For snow versus sun

$$\lambda_{\text{snow}} = \exp(-1.326403 + (-0.3725706(\text{temp_max})) + (-0.2276596(\text{temp_min})) + (33.15028(\text{precipitation})) + (0.1154182(\text{wind})))$$



Fitted Model Prediction Response



$$\lambda_{\text{rain}} = \exp(1.102639 + (-0.1488128 \times 30) + (0.1471095 \times 20) + (33.07871 \times 0) + (-0.3148382 \times 5)) = 0.135$$

$$\lambda_{\text{snow}} = \exp(-1.326403 + (-0.3725706 \times 30) + (-0.2276596 \times 20) + (33.15028 \times 0) + (0.1154182 \times 5)) = 4.496 \times 10^{-8}$$

$$\lambda_{\text{sun}} = 1 - \lambda_{\text{rain}} - \lambda_{\text{snow}} = 0.865$$

```
# Make predictions for a specific example
example_data <- data.frame(
  temp_max = 30,
  temp_min = 20,
  precipitation = 0,
  wind = 5 # Example wind value
)

# Get predicted probabilities for the example data
predicted_probs <- predict(fitted_model, newdata = example_data, type = "prob")

# Convert the probabilities to percentages
predicted_probs_percentage <- predicted_probs * 100

# Format the percentages to avoid scientific notation and round to 2 decimal places
formatted_probs <- format(predicted_probs_percentage, digits = 2, scientific = FALSE)

# Print the predicted probabilities as percentages
print("Predicted Probabilities (in percentage):")
print(formatted_probs)
```

```
[1] "Predicted Probabilities (in percentage):"
      sun      rain      snow
"88.0138894" "11.9861045" " 0.0000061"
```





Analysis



For this weather dataset, the most significant predictor is precipitation, which strongly increases the likelihood of rain. Temperature, specifically the minimum temperature, also plays a key role in predicting snow. Given the complex nature of weather patterns, the model captures the relationship between the predictors and the response variable.

It's important to note that, with the current variability in weather conditions, precipitation and temperature remain the dominant factors influencing weather type.



Thanks for your attention!

