Mount Rainier: The Climbers' Challenge

Authors

- **Subhrajyoty Roy** (MB1911)
- **Abhinandan Dalal** (MB1920)
- **Arindam Roy Chowdhury** (MB1919)
- **Abhishek Bakshi** (MB1938)

What? And Why?

- Mount Rainier: Large active Stratovolcano located 59 miles south south-east of Seattle, Washington.
- **Highest Mountain peak** of state Washington, the Cascade range; and **most topographically prominent** peak of the contiguous United States (ie, leaving out Alaska, Hawaii and islands.)
- Elevation: 4392metres

Mountaineer’s Challenge

Difficult Climbing: Involving traversing largest glacier in the US south of Alaska.

Requires 2 to 3 days of summit, with high failure rates.

Weather and physical conditions posit largest challenges, with the former being very erratic.

Climbing teams require experience in glacier travel, self-rescue, and wilderness travel.

High rising climbers require permits by law.

Our Aims:

Use the Mount Rainier dataset from Kaggle

To model the climbers traffic on various routes based on weather conditions

To model the route wise success proportions based on weather conditions and time of year.

Data description

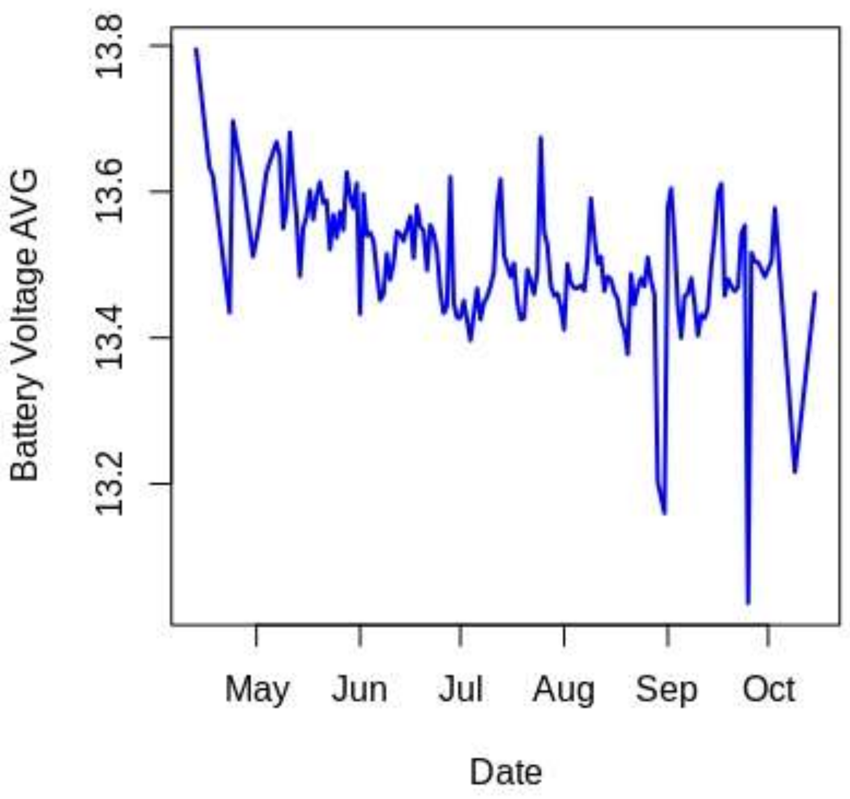
We use the dataset of the Climbier statistics, from **25th of September, 2014; to 27th of November, 2015**.

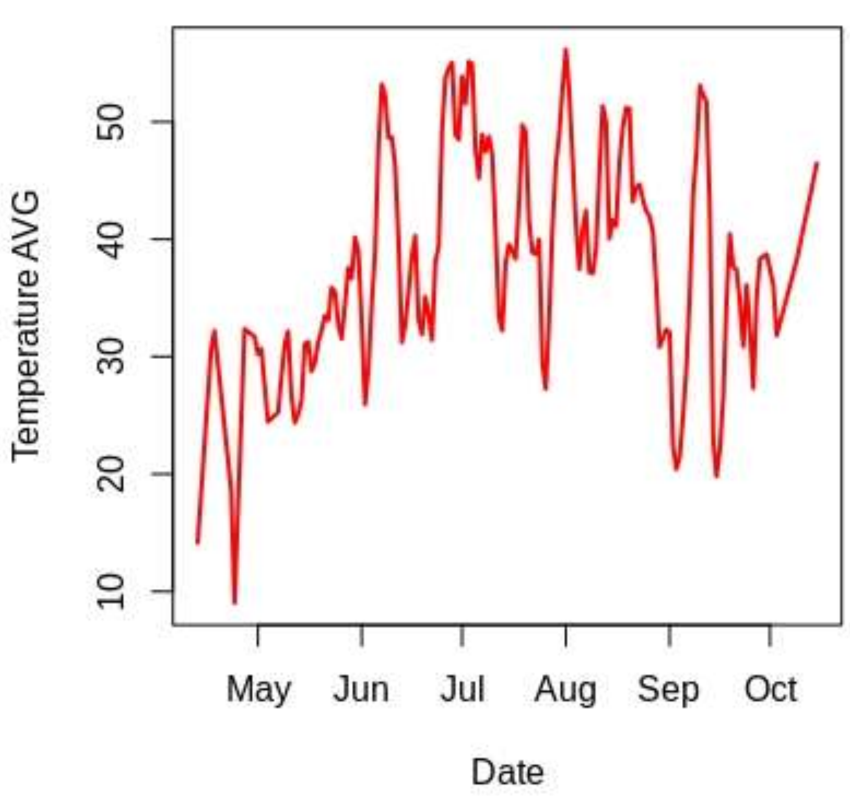
The main routes of climbing are **Disappointment Cleaver** (the default route), **Emmons Winthrop** and **Kautz Glacier**. The remaining routes have too few data points to be considered.

Weather covariates are **battery voltage, average temperature, relative humidity average, average wind speed, average wind direction, solare radiation**, measured each day.

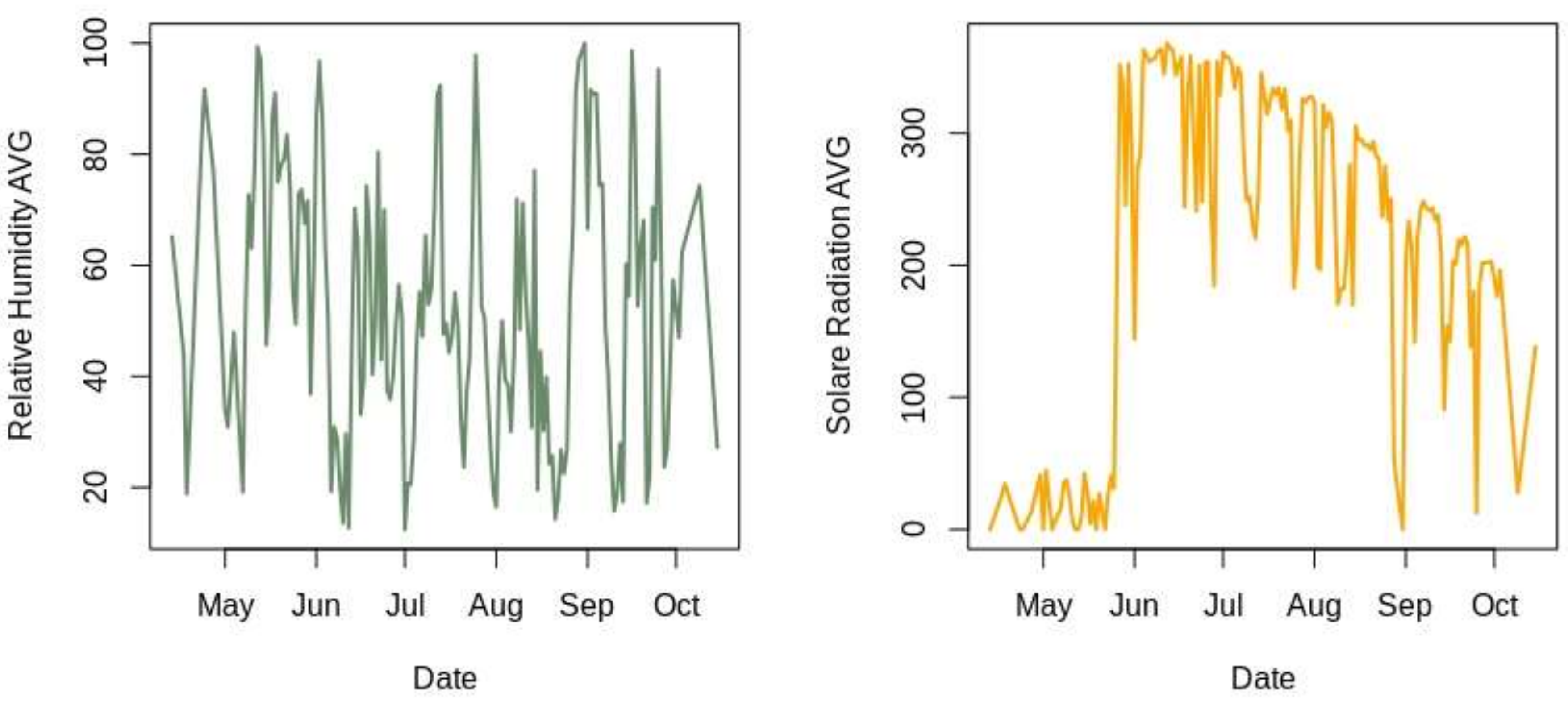
Climbing statistics include **team size**, **day of beginning expedition**, and **number of successes in each team**.

Viewing Weather Conditions

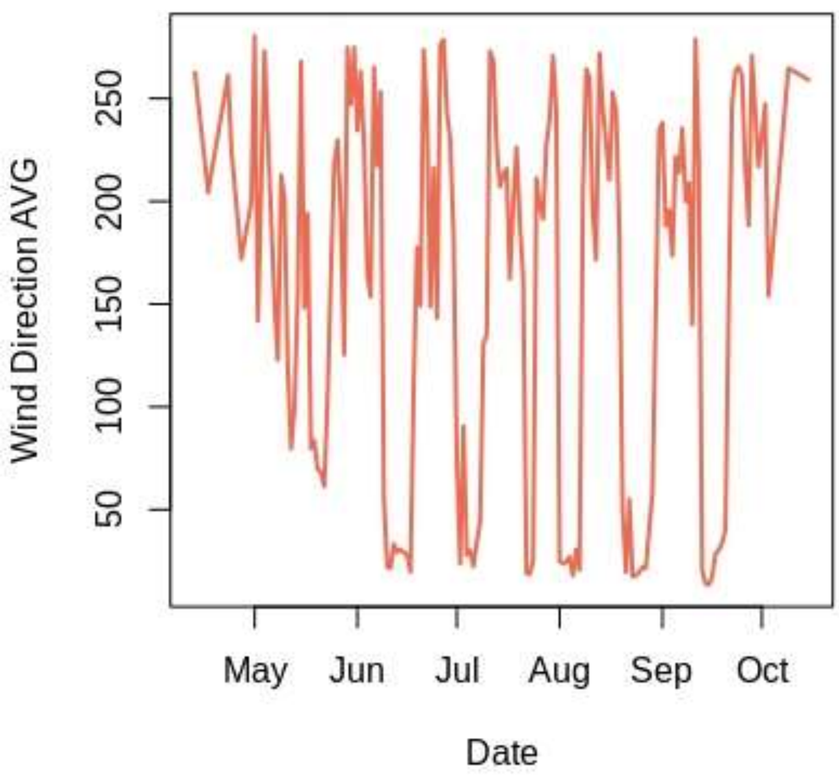


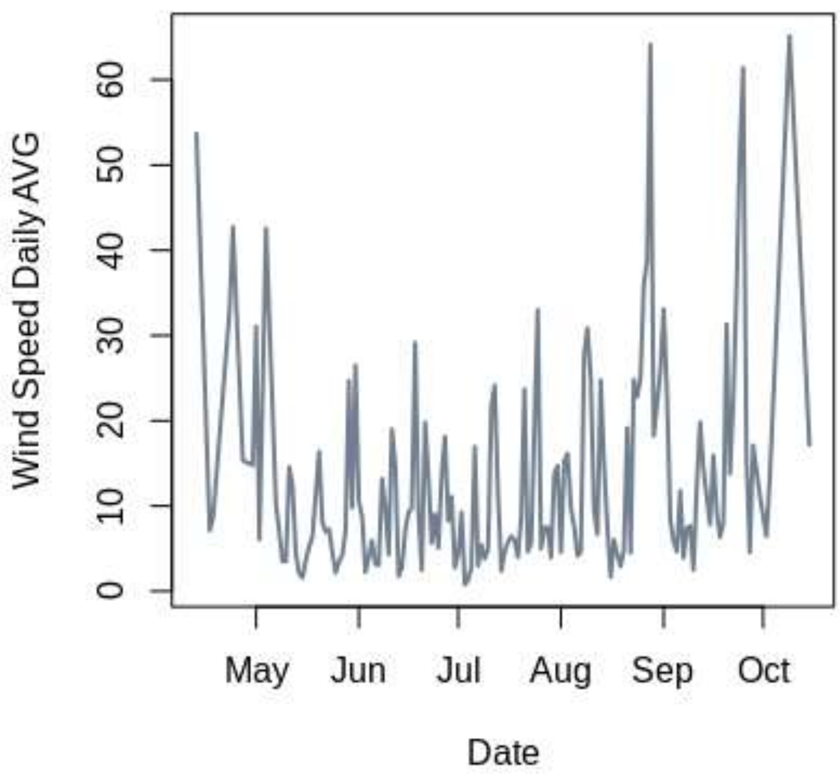


Viewing Weather Conditions



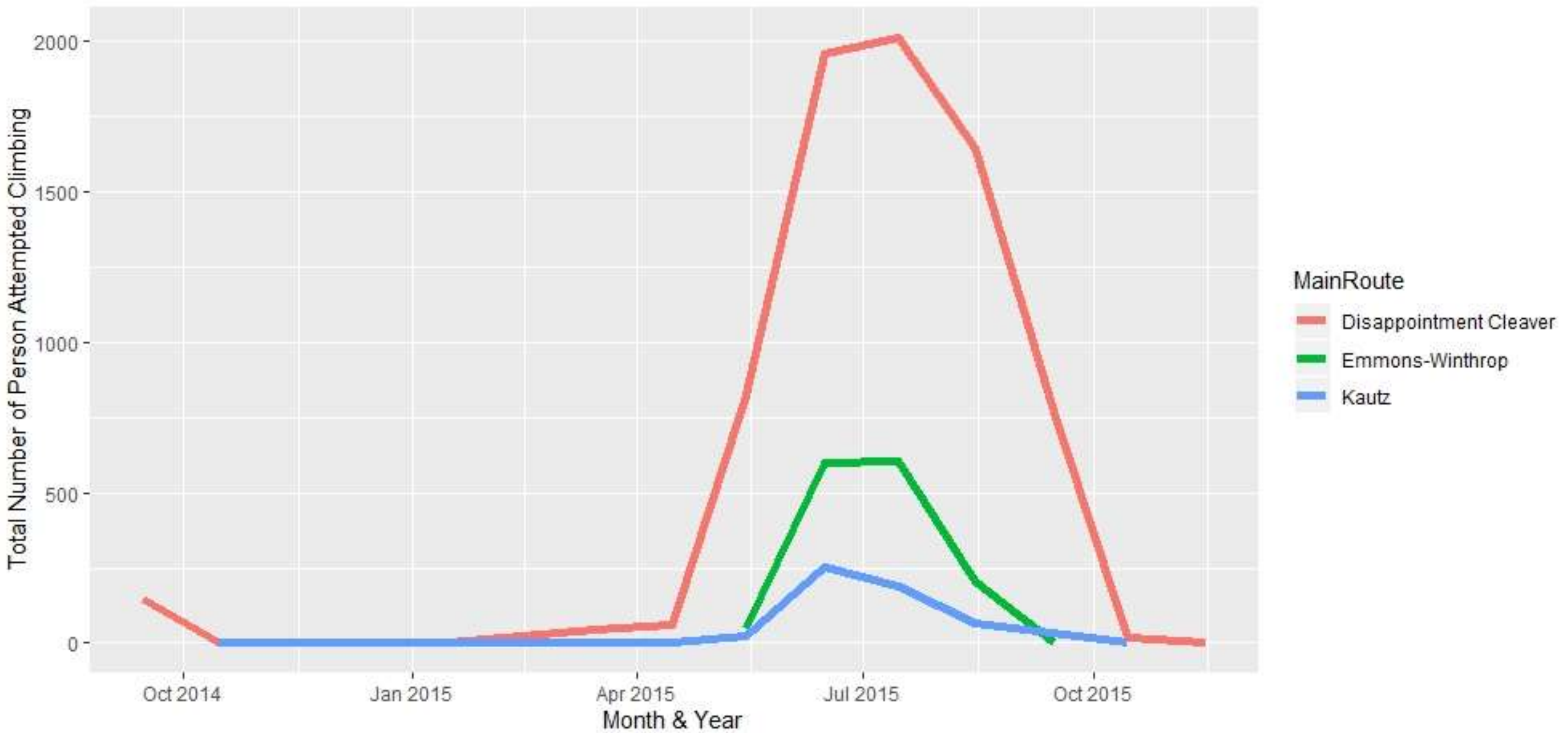
Viewing Weather Conditions



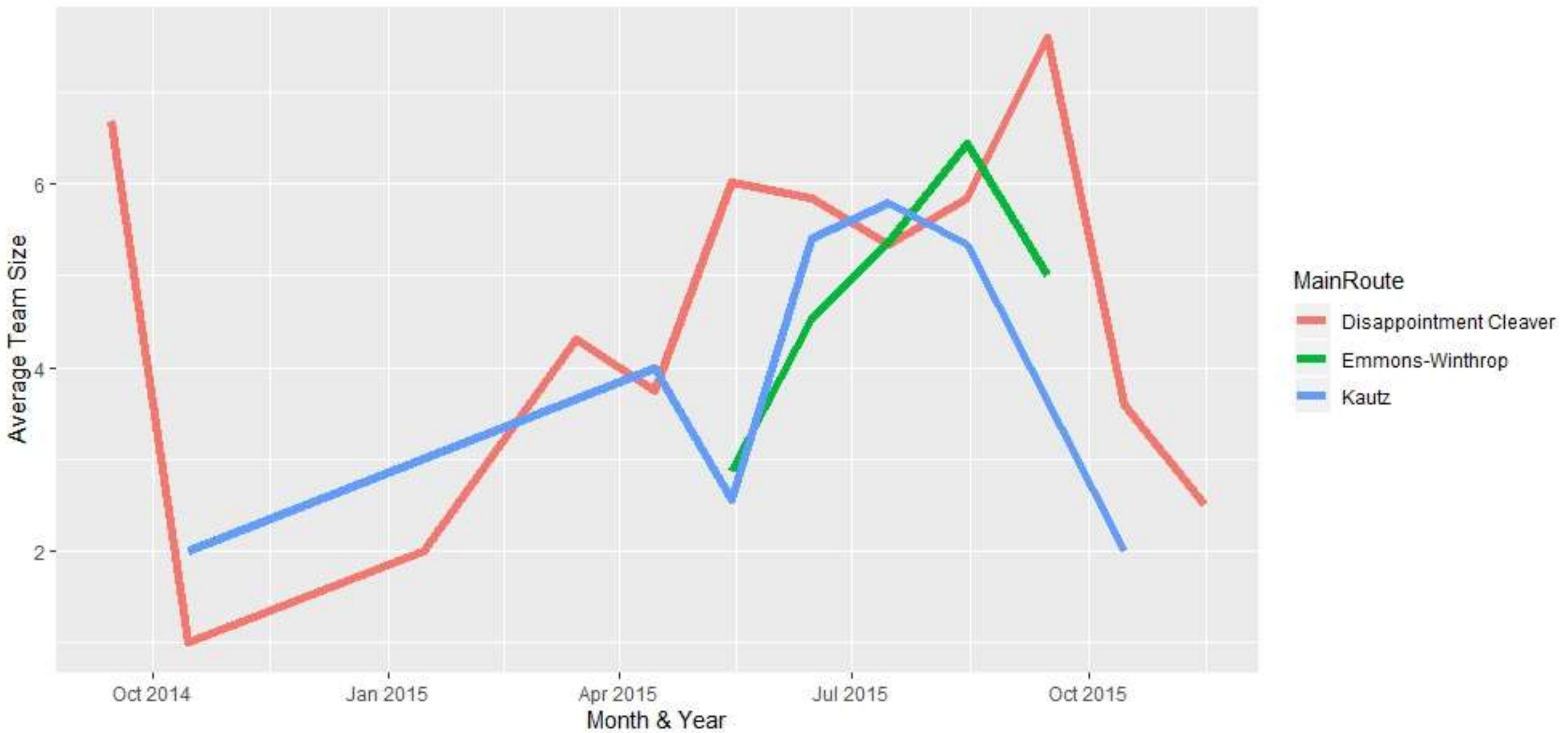


Time Modelling

Time Modelling



Time Modelling



Time Modelling

- Note that the graphs show that the number of attempts upsurge during the middle of the year. This trend is seen in both years of the dataset.
- Hence we use a sinusoidal time model, consisting of sine and cosine components with the crest at the middle of the year.

$$\gamma_1 \sin\left(\frac{\pi}{365}(t-t_0)\right)+\gamma_2 \cos\left(\frac{\pi}{365}(t-t_0)\right)$$

Month	Teams	Team Size
09/14	22	6.68
10/14	3	1.33
01/15	1	2
03/15	10	4.3
04/15	17	3.76
05/15	163	5.50
06/15	514	5.47
07/15	522	5.38
08/15	325	5.89
09/15	102	7.58
10/15	6	3.33
11/15	2	2.5

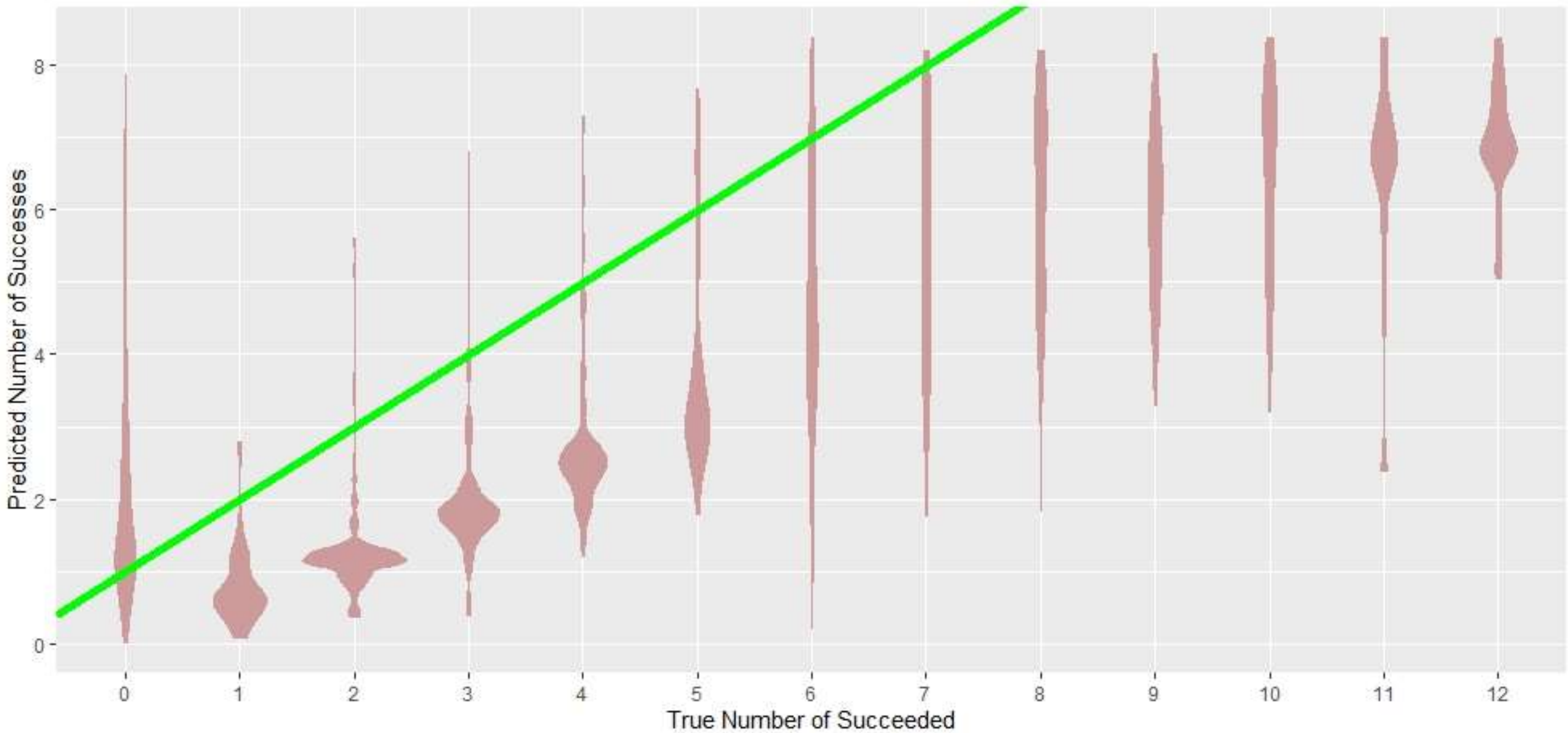
Model 1: Logistic Regression

- We begin with the standard tool: Using a logistic model to predict the successes and failures.

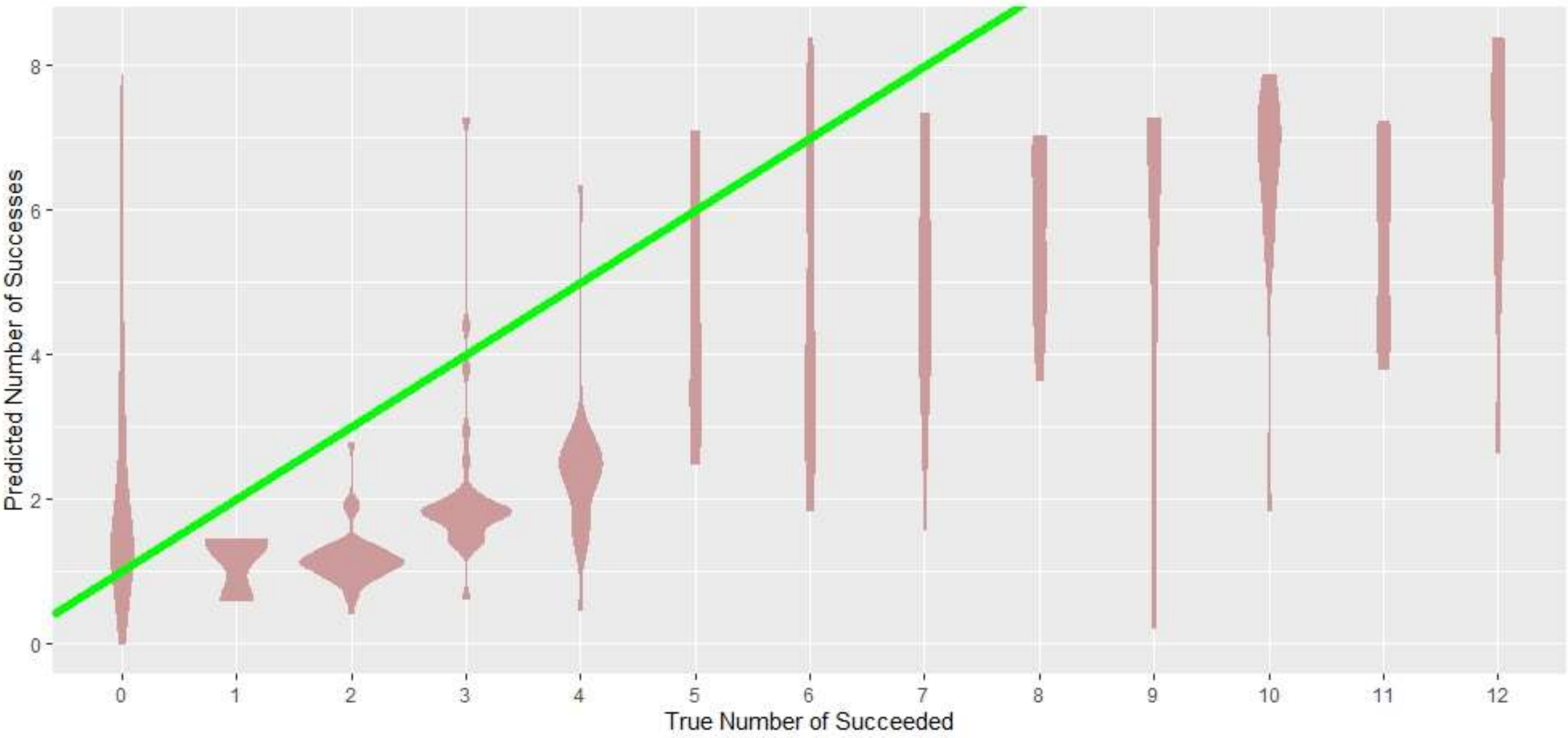
$$\log\left(\frac{p_i}{1-p_i}\right)=\alpha_0+\alpha_j+\sum_{k=1}^{p}\beta_k x_{ik}+\gamma_1\sin\left(\frac{\pi t}{365}\right)+\gamma_2\cos\left(\frac{\pi t}{365}\right)$$

- *j*=1, 2 or 3 according as the *j*th route is traversed, *x* are the weather covariates as observed, and *t* is the traverse date of the year.

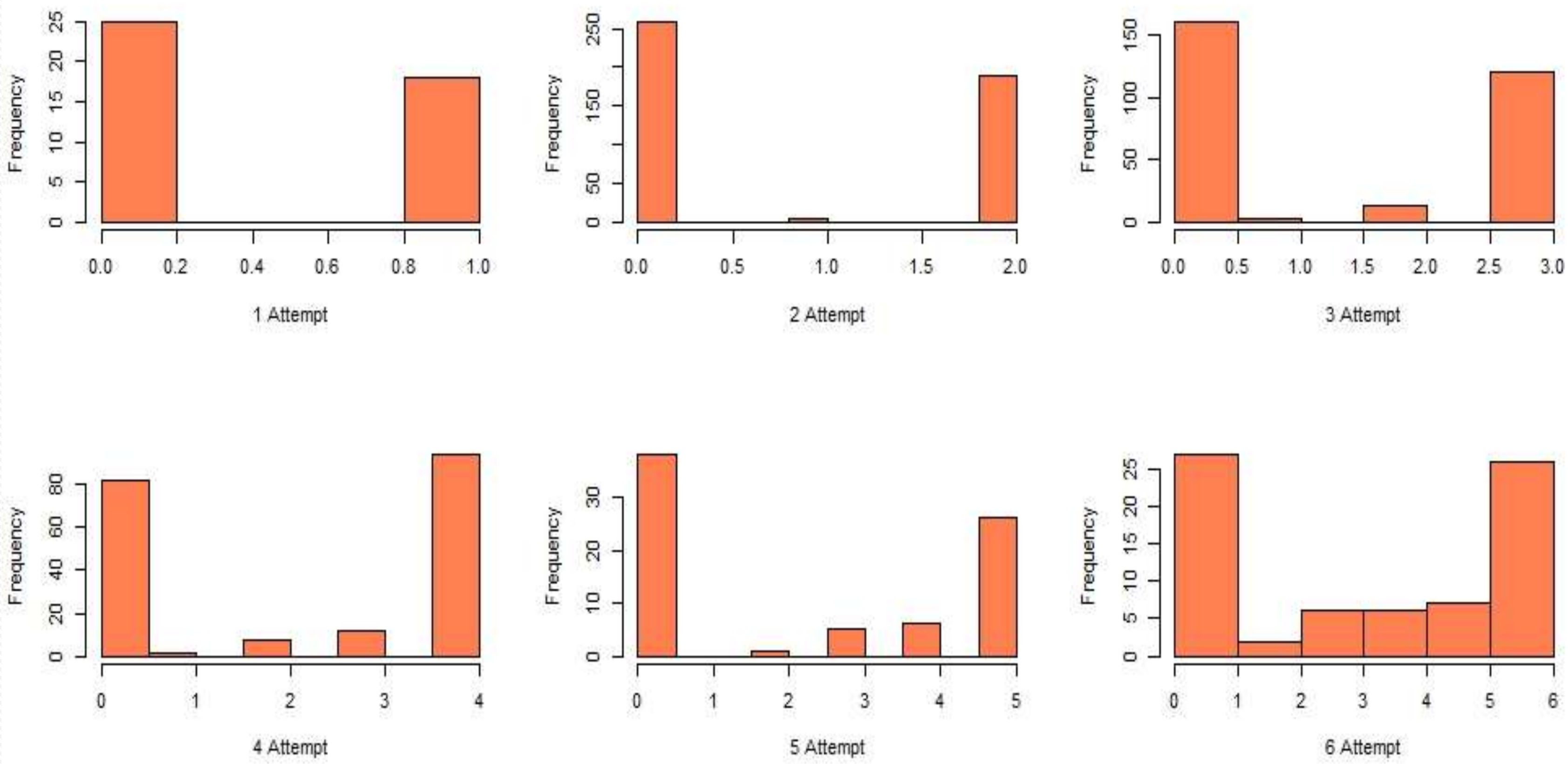
Fit from Logistic Regression: Training



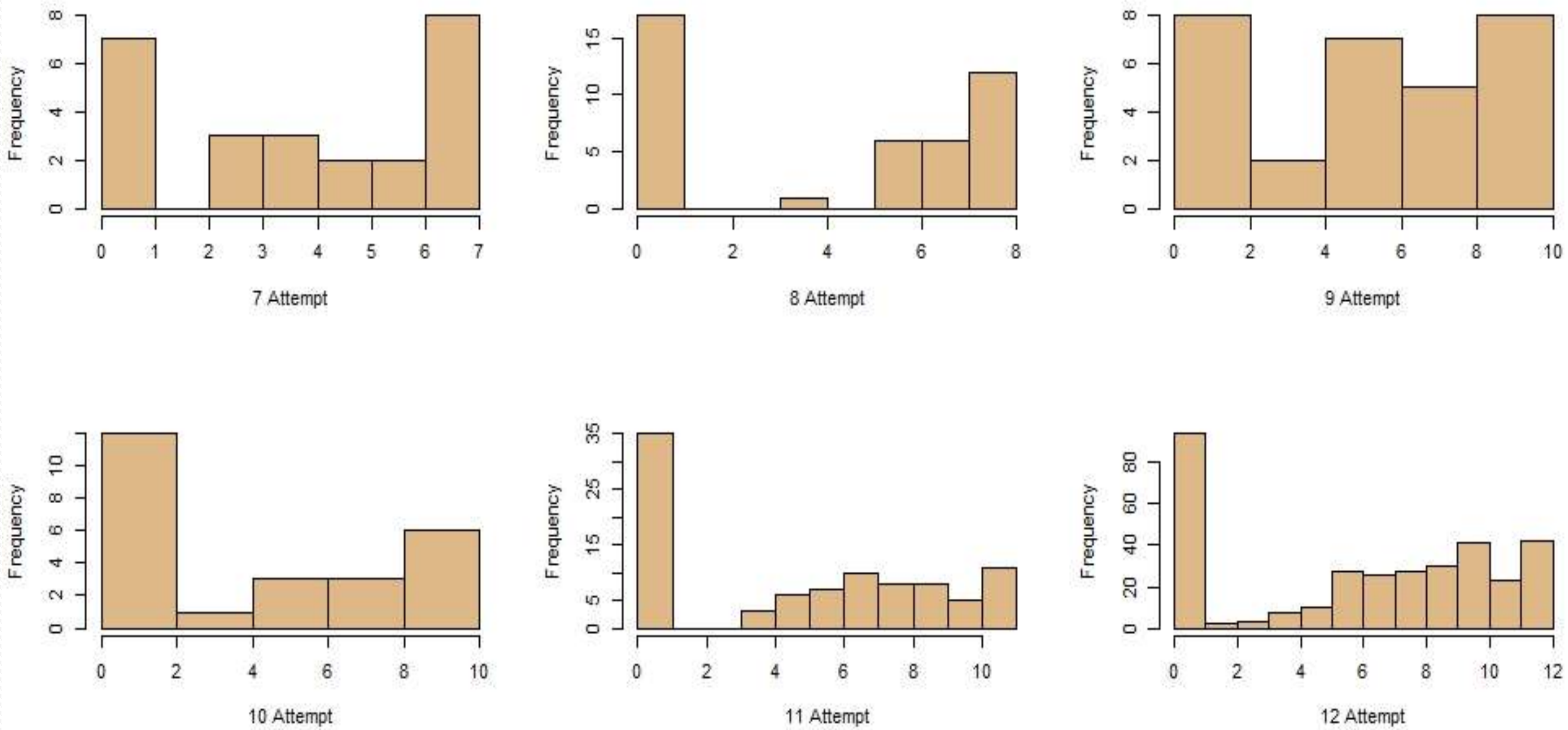
Fit from Logistic Regression: Test



Histograms of success frequencies



Histograms of success frequencies



Why did Logit regression fail?

Notice the peaks in the success proportions, mostly at 0 and n, n being the team size .

Why does this happen?

In a team mountaineering expedition, your success is not independent of your teammates

If your teammates fail, it is very unlikely that you succeed alone

If all your teammates succeed, it is highly likely that you succeed too.

Model 2: Independent ZIPs

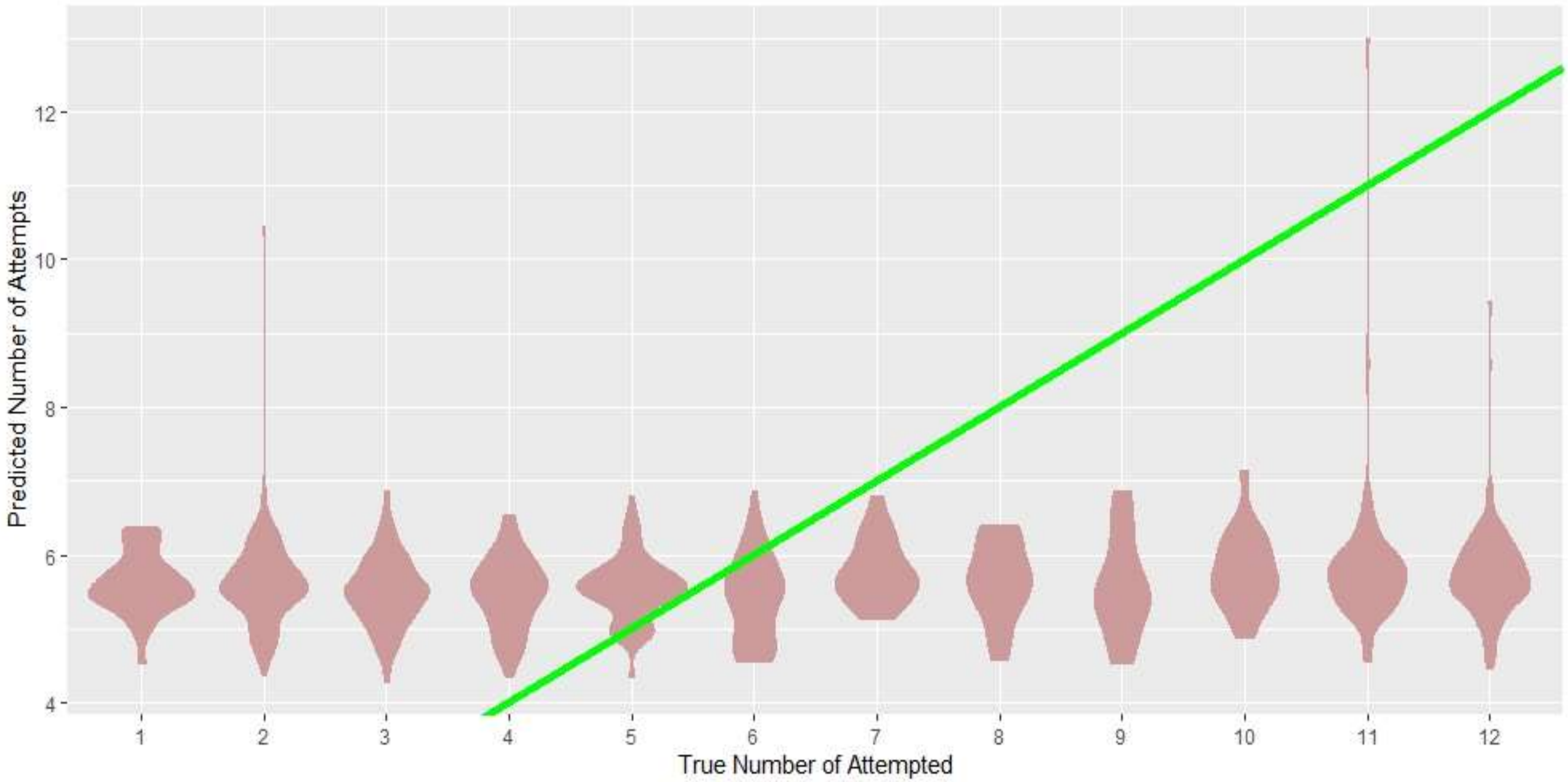
Rising from the ashes of the previous model, we propose an alternative one.

If our team size had follows a Poisson distribution, and the successes arrivals and failure arrivals occur independently, we would have had the Binomial distribution.

Due to the high number of peaks at the end, it looks like both the successes and failures are 0 inflated.

We try with independent **0 inflated successes** and **0 inflated failure Poisson model** (assuming arrivals are Poisson).

Fits of Independent ZIPs: Training Set

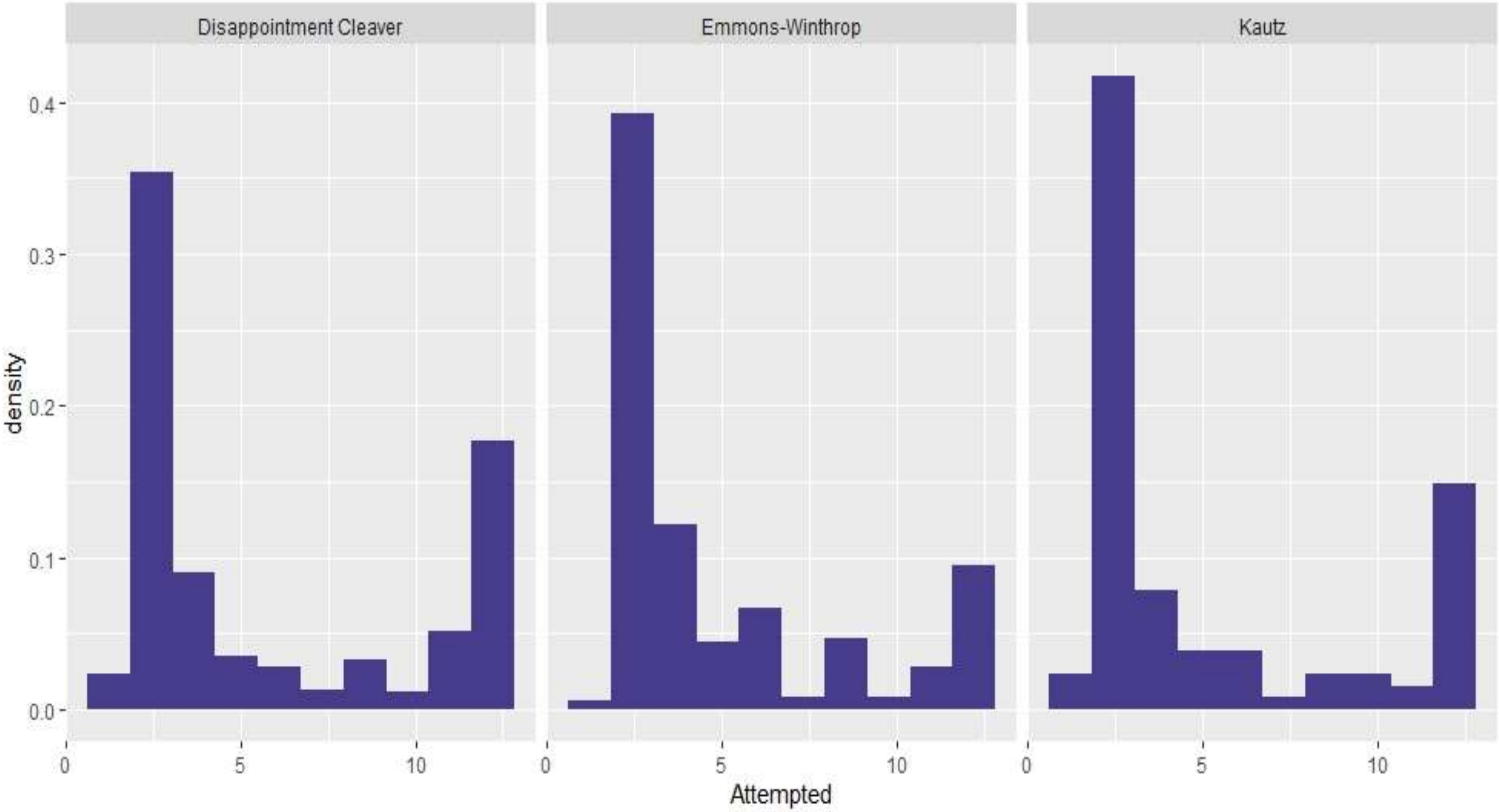


Assessing failure of Model 2

We used the number of attempts to be the sum of the number of successes and number of failures.

The failure of the ZIP model lies in the failure of the log-linear model fits for the two Poissons.

Histograms of Attempts



Model 3: Separating N and p

- Till now, we had tried to model the number of successes directly.
- However, the number of successes depend on N, and hence we try to model team size N, and the number of successes conditioned on N.
- Modelling N: From the histogram of attempts, the mutimodality in the distribution of team size is prominent.
- Hence, we use a three component mixture of Poisson distributions to model the team size N.

Modelling N: Mixture of Poisson

$$N = N_1 X_1 + N_2 X_2 + N_3 X_3$$

- where N is the team size, N_i s are independent Poi(λ_i), i=1(1)3; (X_1,X_2,X_3)~Multinomial(1,p_1,p_2,p_3)
- The Poisson means are modelled by log-linear models of the covariates, and the multinomial proportions by a multinomial logistic model.

Results from Poisson mixture: Components

	Component 1	Component 2	Component 3
Intercept	**Estimate: 1.045** **p value< 2e-16**	**Estimate: 2.3798** **p value< 2.2e-16**	**Estimate: 1.104** **p value< 2e-16**
Route Emmons-Winthrop	Estimate: 0.1519 p value: 0.125	Estimate: -1.1465 p value: 0.174	**Estimate: 1.187** **p value< 2.2e-16**
Route Kautz	Estimate: 0.0225 p value: 0.892	Estimate: -1.272 p value: 1.051e-08	**Estimate: 1.278** **p value< 2.2e-16**

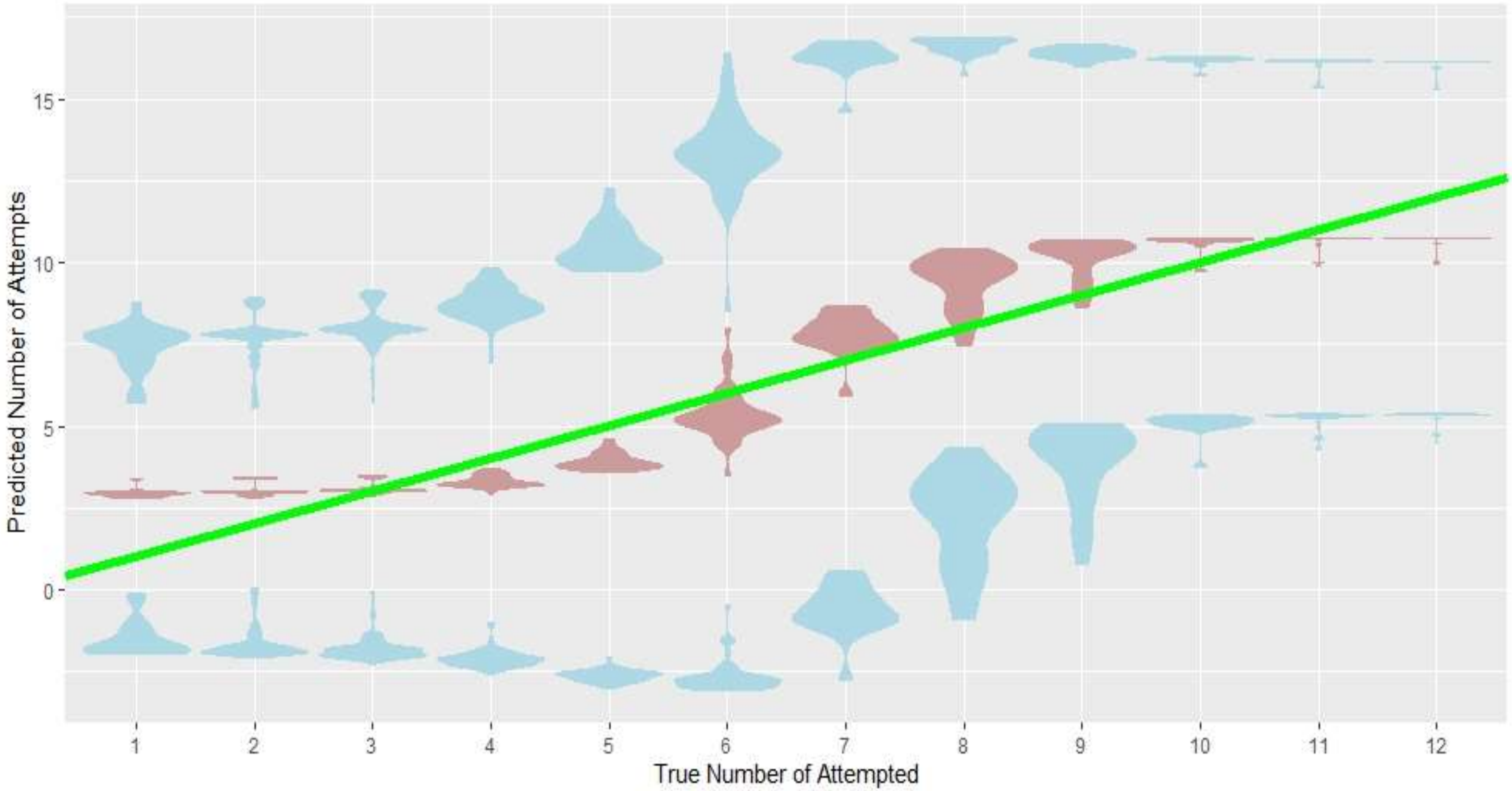
The weather covariates turn out to be not significant in the log-linear means of any component. We use the Disappointment Cleaver route as the base.

Results from Poisson mixture: Mixture probs.

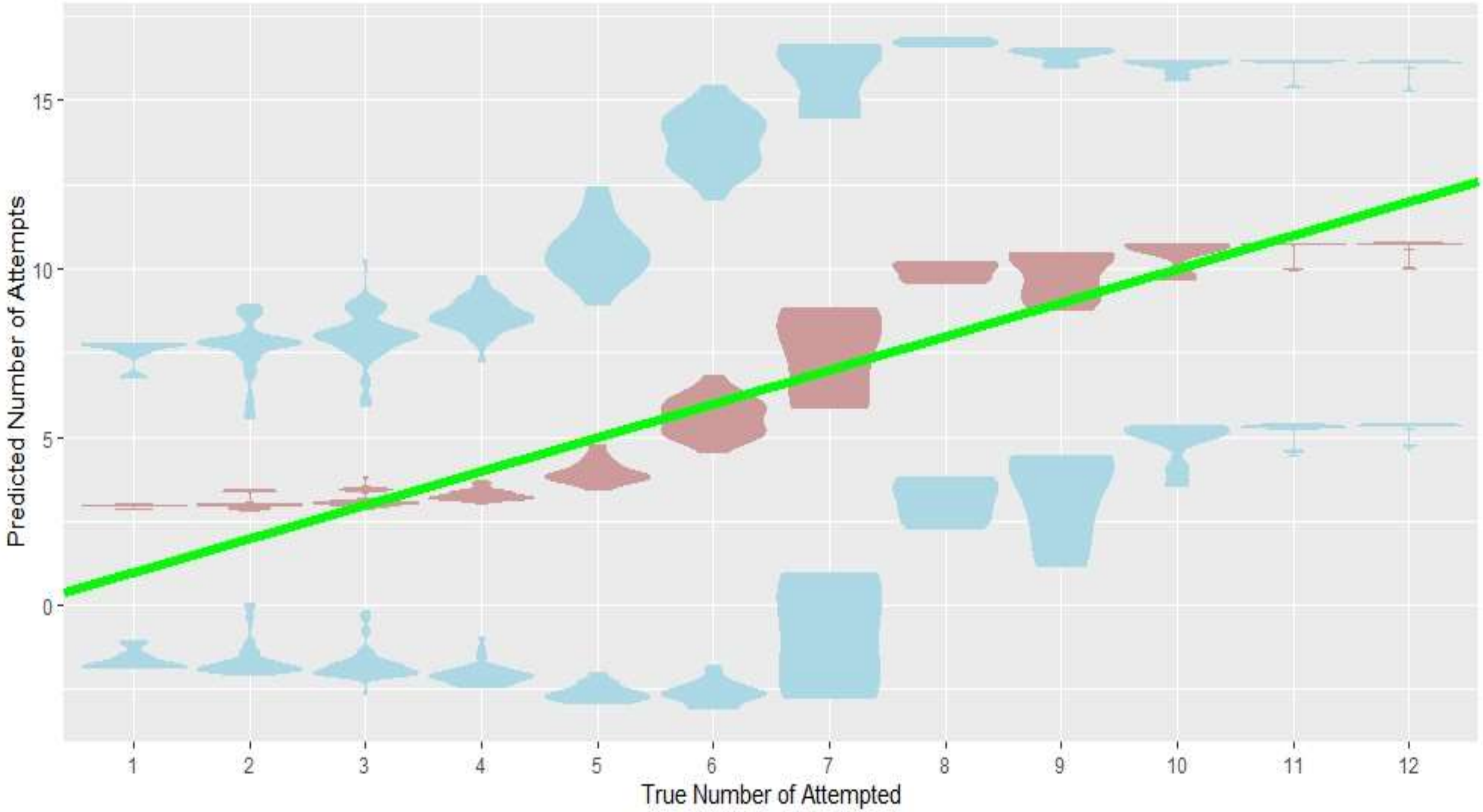
	Mix Prob 2		Mix Prob 3	
	Esti	pval	Esti	pval
Intercept	-30.01	0.65	-110	0.122
Route Emons	**-8.315**	**0.005**	**-5.429**	**0.040**
Route Kautz	-4.309	0.636	-4.739	0.249
sin(pi*t/365)	**23.77**	**0.027**	**3.422**	**0.000**
cos(pi*t/365)	**-3.72**	**0.01**	**-4.964**	**0.000**
Battery Voltage	0.82	0.86	6.169	0.226
Temperature	-0.02	0.74	-1.238	0.822
Rel.Humidity	-0.005	0.76	-2e-02	0.249
Wind speed	0.038	0.158	1.618	0.519
Wind direction	-0.001	0.58	-4.337	0.837
Solar radiation	0.001	0.813	-1.399	0.807

The weather covariates turn out to be not significant in any mixture component as well. We use the Disappointment Cleaver route as the base, and component 1 to be the reference.

Fit of Number of Attempts: Training



Fit of Number of Attempts: Test



Modelling Successes given Team Size

To accommodate the peaks in the 0 and N, we use a 0 and N inflated binomial model (ZNIB)

The 0 and N inflated binomial distribution arises when two zero-inflated Poisson count processes are constrained by their sum total.

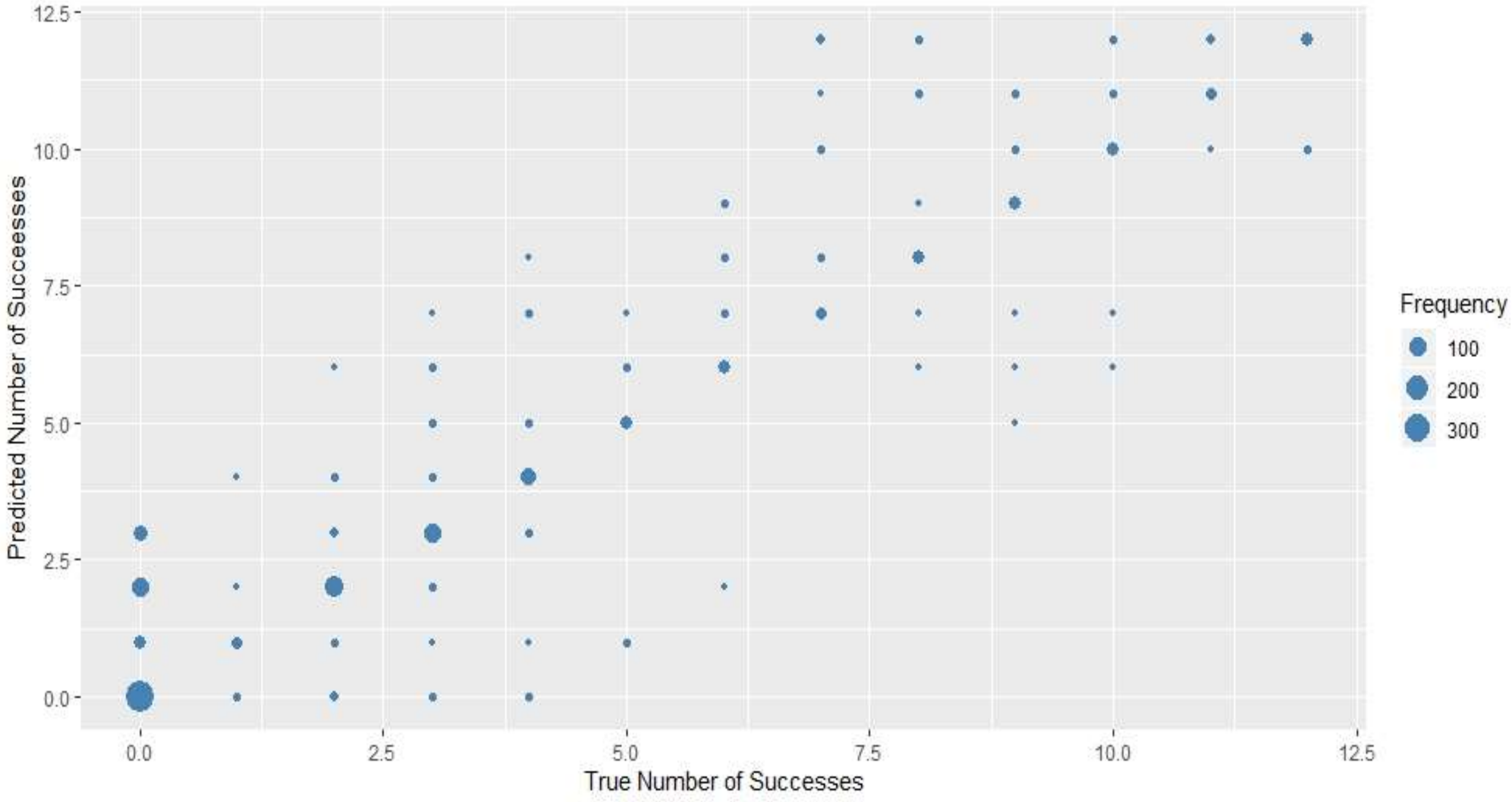
Thus we keep the idea of Model 2, but now, we model these mixtures probabilities using multinomial logistic regression.

No. of successes $X|N \sim$ 0 wp p_1, Y wp p_2, N wp p_3, where 0 and N are degenerate, and Y is Bin(N,*p*).

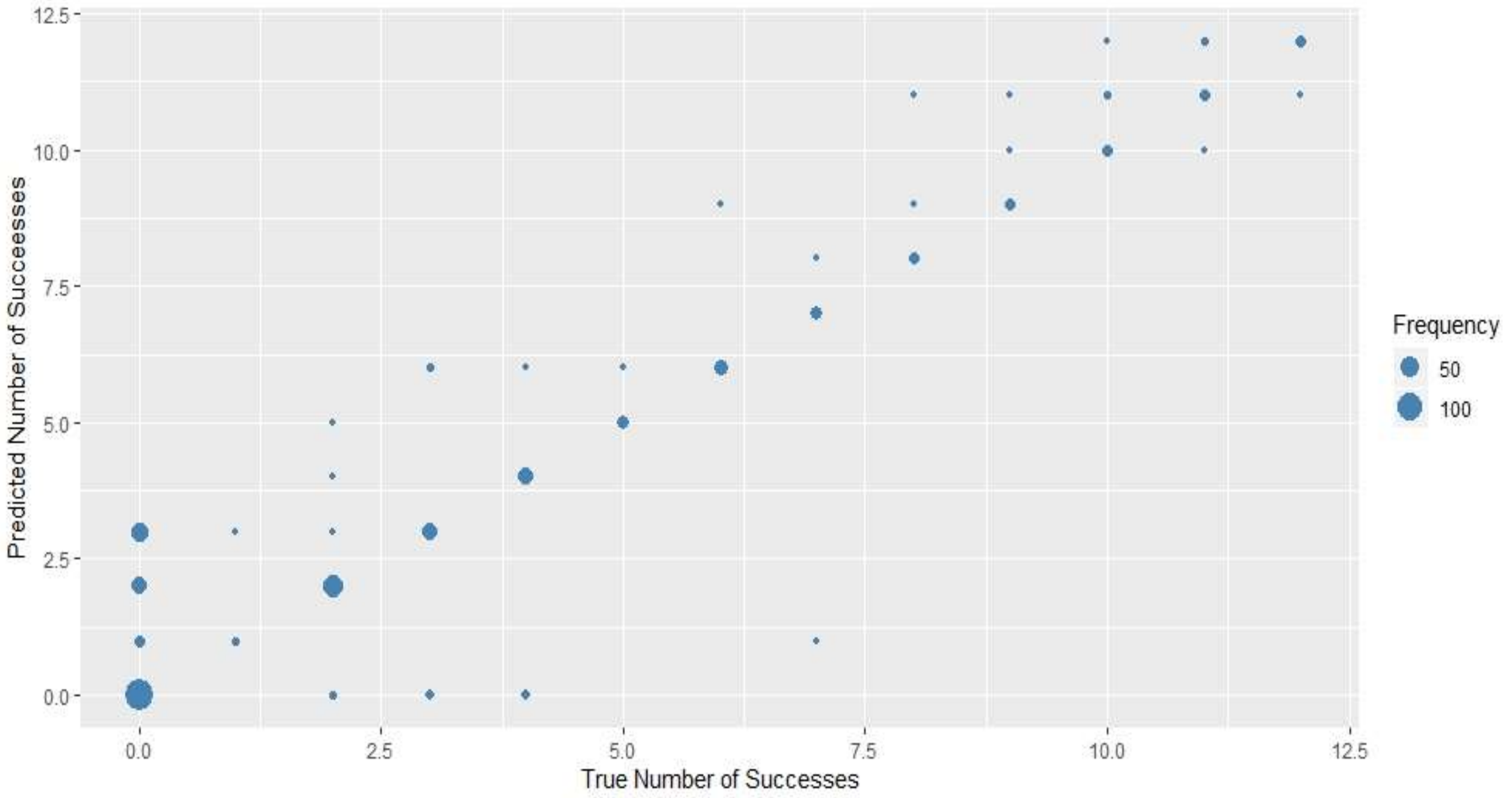
Results show weather covariates to be finally significant

```
> test.inflbinom(form, op)
$`Zero Inflated Model`
                      Estimate   Std.Error      Z-value       p-value Signif.codes
1               (Intercept)  1.7492614 2.946012729    0.5937725  5.526643e-01
2   MainRouteEmmons-Winthrop  0.6385243 5.756641369    0.1109196  9.116801e-01
3            MainRouteKautz  0.8049861 6.126685992    0.1313901  8.954667e-01
4   sin((pi/365) * TransDate)  0.7317293 0.033206437   22.0357658 1.308145e-107          ***
5   cos((pi/365) * TransDate) -0.9318138 0.014570052  -63.9540457  0.000000e+00          ***
6      `Battery Voltage AVG`  1.6937292 0.032154539   52.6746523  0.000000e+00          ***
7        `Temperature AVG` -0.3320500 0.003748464  -88.5829286  0.000000e+00          ***
8    `Relative Humidity AVG`  0.6827024 0.001254195  544.3350630  0.000000e+00          ***
9     `Wind Speed Daily AVG`  1.9828078 0.005330472  371.9760351  0.000000e+00          ***
10     `Wind Direction AVG` -0.9512195 0.001539230 -617.9838997  0.000000e+00          ***
11    `Solare Radiation AVG` -2.2523472 7.087719225   -0.3177817  7.506506e-01

$`Binomial Model`
                       Estimate    Std.Error     Z-value       p-value Signif.codes
1               (Intercept)  1.39199143 3.0938462088   0.4499226  6.527662e-01
2   MainRouteEmmons-Winthrop  0.78767494 0.7420851648   1.0614347  2.884924e-01
3            MainRouteKautz  2.13636687 1.1998085626   1.7805898  7.497949e-02            .
4   sin((pi/365) * TransDate)  0.76954260 0.1037894278   7.4144604  1.221210e-13          ***
5   cos((pi/365) * TransDate)  1.51326844 0.0187282389  80.8014276  0.000000e+00          ***
6      `Battery Voltage AVG` -0.59113941 0.0179221318 -32.9837666 1.388324e-238          ***
7        `Temperature AVG` -0.00209541 0.0014524751  -1.4426479  1.491197e-01
8    `Relative Humidity AVG` -0.23103451 0.0271976469  -8.4946508  1.985289e-17          ***
9     `Wind Speed Daily AVG` -0.05396779 0.0035389977 -15.2494558  1.660111e-52          ***
10     `Wind Direction AVG` -0.10548299 0.0039989195 -26.3778735 2.458780e-153          ***
11    `Solare Radiation AVG`  0.12323413 0.0002417459 509.7672681  0.000000e+00          ***

$`End Inflated Model`
                        Estimate    Std.Error      Z-value       p-value Signif.codes
1               (Intercept) -0.367064699 1.490264e+01  -0.02463085  9.803494e-01
2   MainRouteEmmons-Winthrop  0.343897123 1.080526e-01   3.18268203  1.459178e-03           **
3            MainRouteKautz  1.092544381 1.692586e-01   6.45488395  1.083019e-10          ***
4   sin((pi/365) * TransDate)  4.970575941 1.211009e+00   4.10449115  4.052058e-05          ***
5   cos((pi/365) * TransDate)  0.993955617 2.031283e-01   4.89324138  9.918863e-07          ***
6      `Battery Voltage AVG`  0.079220067 1.066767e+00   0.07426184  9.408020e-01
7        `Temperature AVG`  0.305456373 1.240726e-02  24.61916111 7.876598e-134          ***
8    `Relative Humidity AVG`  0.115973405 4.506230e-03  25.73623851 4.596254e-146          ***
9     `Wind Speed Daily AVG` -0.277985988 1.335037e-02 -20.82233784  2.716058e-96          ***
10     `Wind Direction AVG`  0.006961343 5.546459e-04  12.55096780  3.926810e-36          ***
11    `Solare Radiation AVG`  0.002259176 5.792641e-04   3.90007879  9.616139e-05          ***

>
```

Fits of Number of Successes: Training



Fits of Number of Successes: Test



Comparing All models

Model 1: Logit	Model 2: Independent ZIP	Model 3: Mix Poisson and ZNIB
AIC: 6857.4	AIC: 6374.770	AIC: 6170

- AIC is less in Model 3, hence Model 3 has an advantage.
- However, with respect to the prediction accuracy, it is fairly evident that Model 3 is a clear winner. Hence we declare our final model to be the team size to be a mixture of Poisson distribution, and number of successes given team size to have ZNIB.

Remarks

- Note that, in our final proposed model, the team size does not depend on the weather covariates.
- Giving it a thought, note that the team is decided apriori, and hence the individual that day weather is not factored in deciding teams.
- However, the time of the year is significant in team size. Thus an overall idea of seasons factors in deciding team size traffic.
- The team size distribution would help the climbing industry to provide better service to climbers.

Remarks

- However, the success proportions do depend on the weather covariates.
- This dependence is significant not only in the 0 and end components, but also in the Binomial component of the success proportions.
- The success of our endeavour lies in the overall satisfactory prediction of success of the climbing exercise, which had been a challenge due to the unpredictability of Mount Rainier route and weather conditions.

Return to Base Camp

THANK YOU