

IoT Analytics

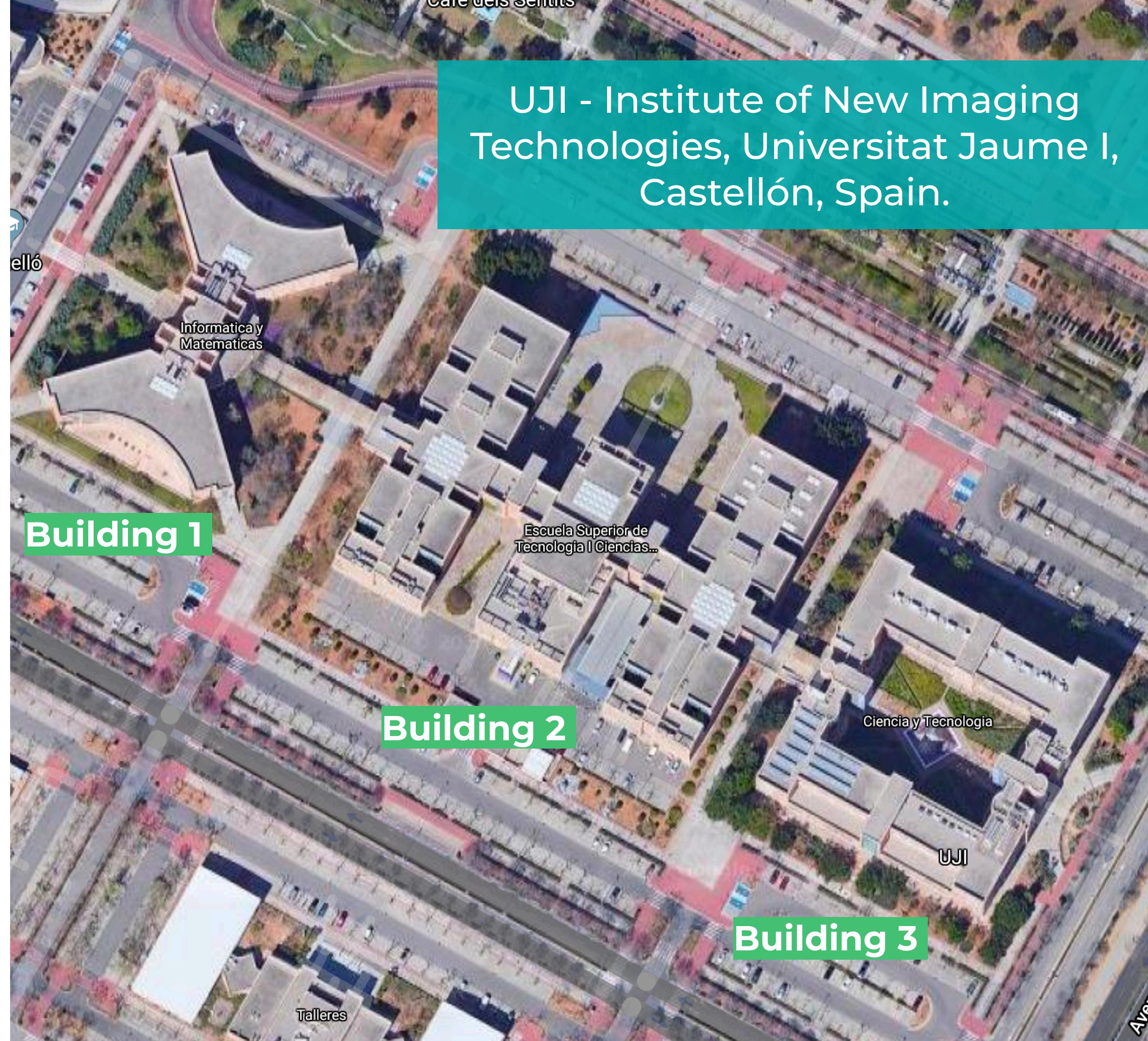
Indoor Locationing Wifi Fingerprint

by Matias Barra

DATASET

Accurate Indoor Positioning

Predict:
Building
Floor
Longitude
Latitude

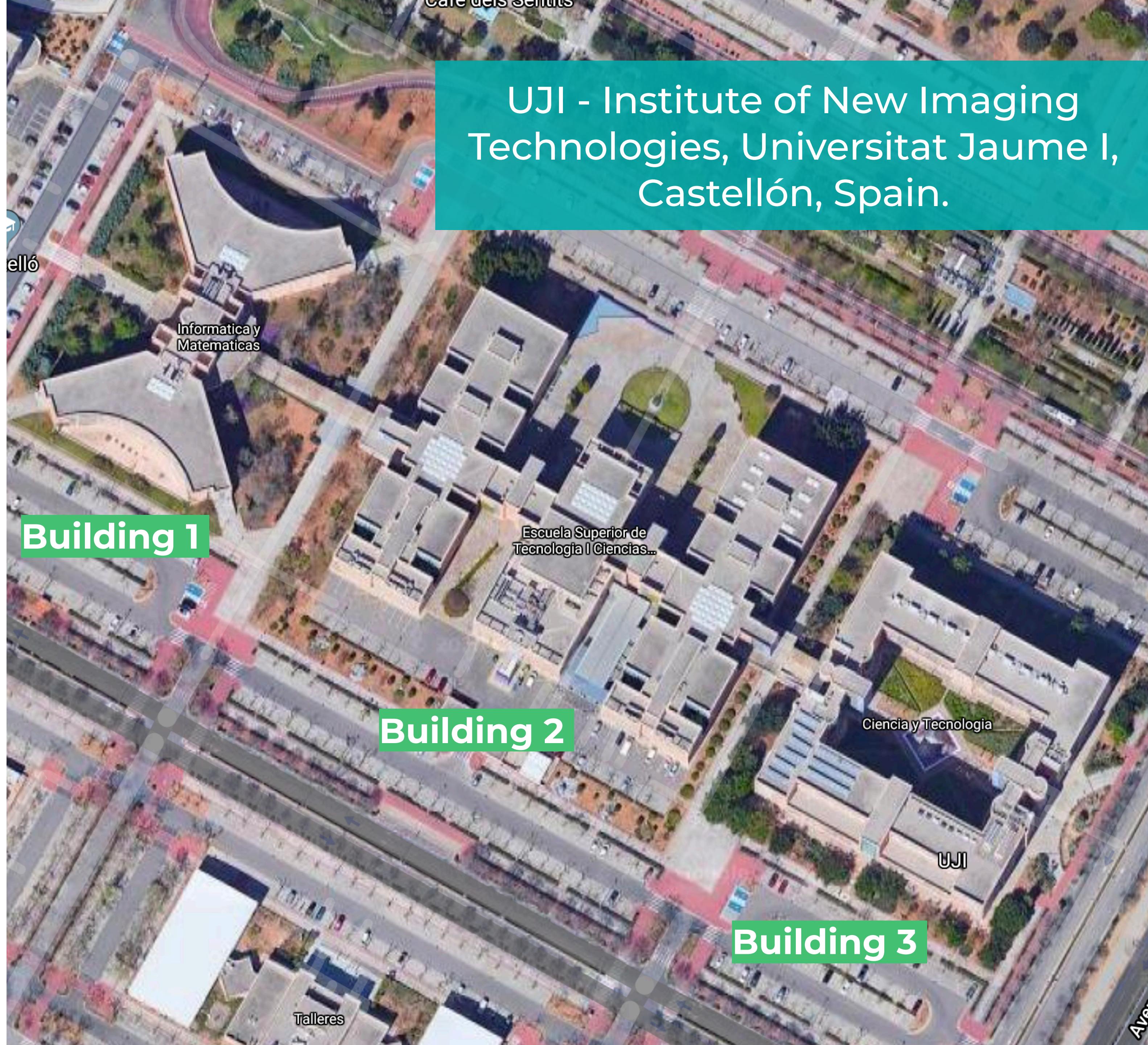


**UJI - Institute of New Imaging
Technologies, Universitat Jaume I,
Castellón, Spain.**

PROBLEM

Accurate Indoor Positioning

Predict:
Building
Floor
Longitude
Latitude



DATA EXPLORATION

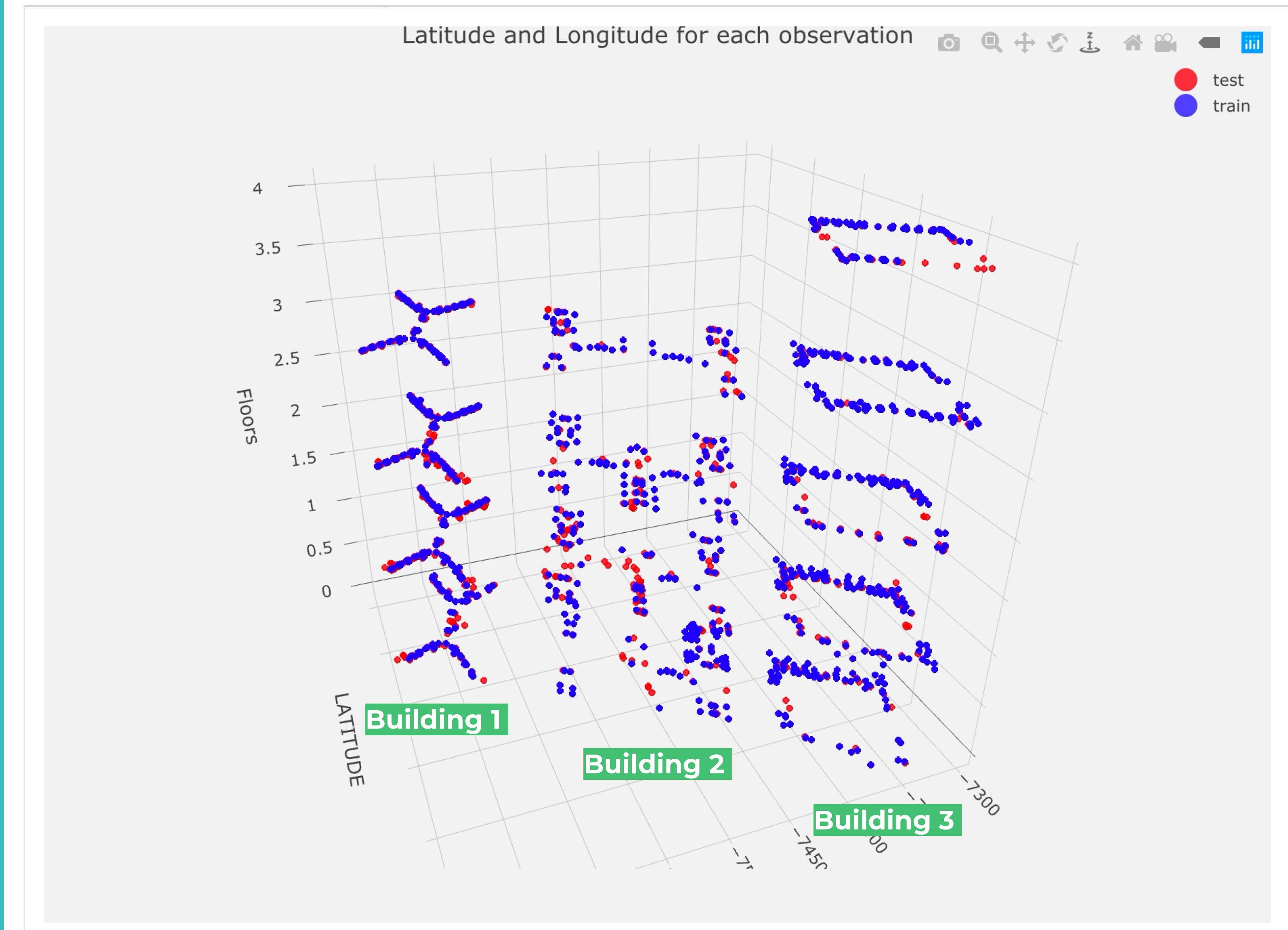
Sites might present problems:

Building 1 – Floors: 0 and 1.
In the middle area and two corners. In general is ok.

Building 2 – Floors: 0, 1, 3.
In the middle area.

Building 3 – Floor: 4.
Lower right corner has no measurements.

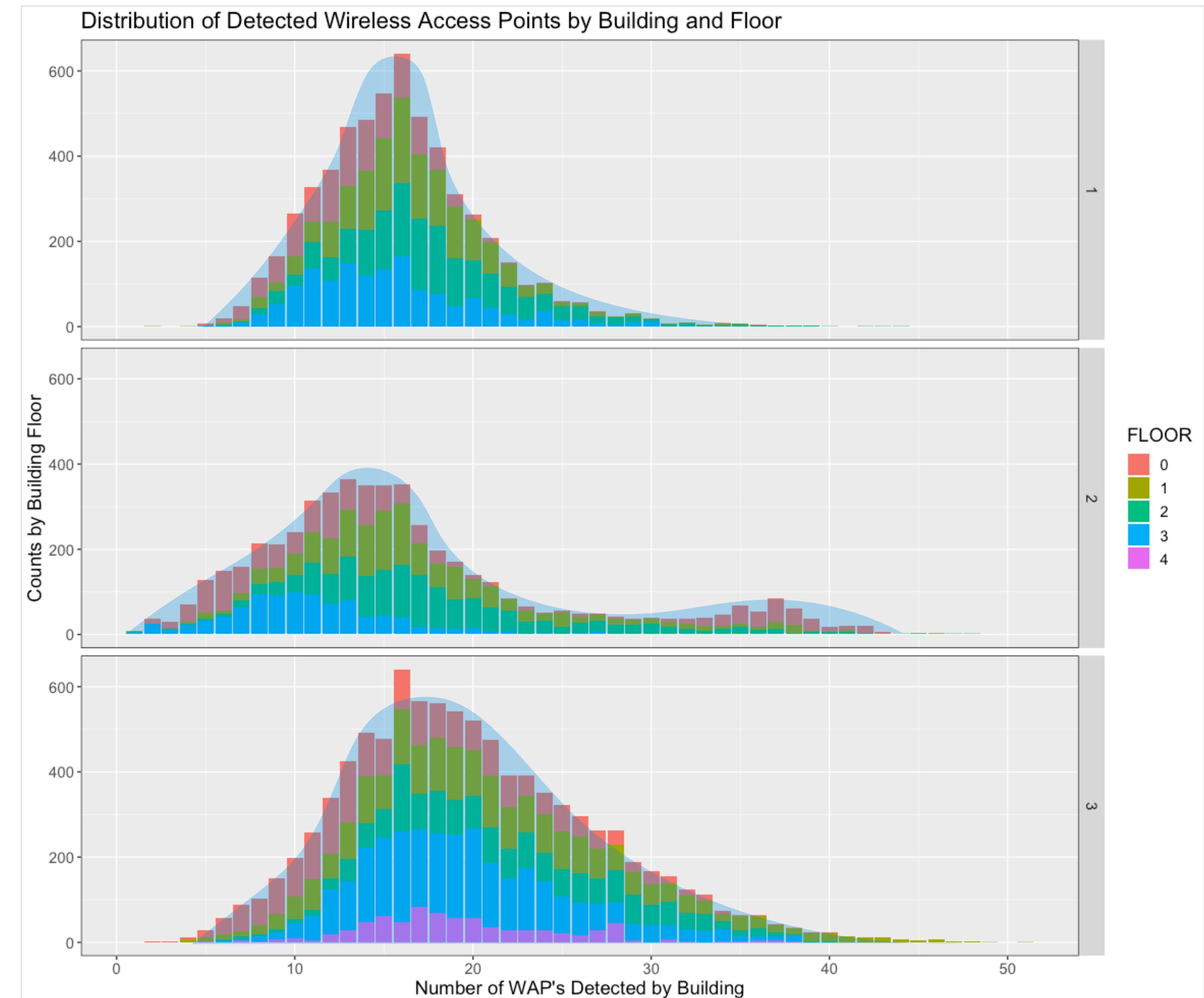
Distribution of each observation by building and floor
Train and test datasets



DATA EXPLORATION

The distribution of WAPs detected in **building 2** is more spread than others two.

Building 3 is the only one with a 4th floor and it also has spikes in WAPs detected at **17** and **28**.



FEATURE SELECTION

- Removing unnecessary columns
- Removing rows with all NAs values
- The highest WAP
- Convert NA's to -110 numeric value

Train Data Set

520 W-Fi Access Points detected.
19.277 records used for training
(fingerprints)

Test Data Set

520 W-Fi Access Points detected, 1.111
records used for validation.

BUILDING PREDICTIONS

Building ID		Confusion Matrix				
Support Vector Machine [svm]	Kernel: Linear		SVM	Reference		
	Prediction	1		2	3	
	1	536		1	0	
	2	0		306	0	
K-Nearest Neighbor [K-NN]	3	0		0	268	
	K = 5		K-NN	Reference		
	Prediction	1		2	3	
	1	534		1	0	
	2	2		307	1	
	3	0		0	267	

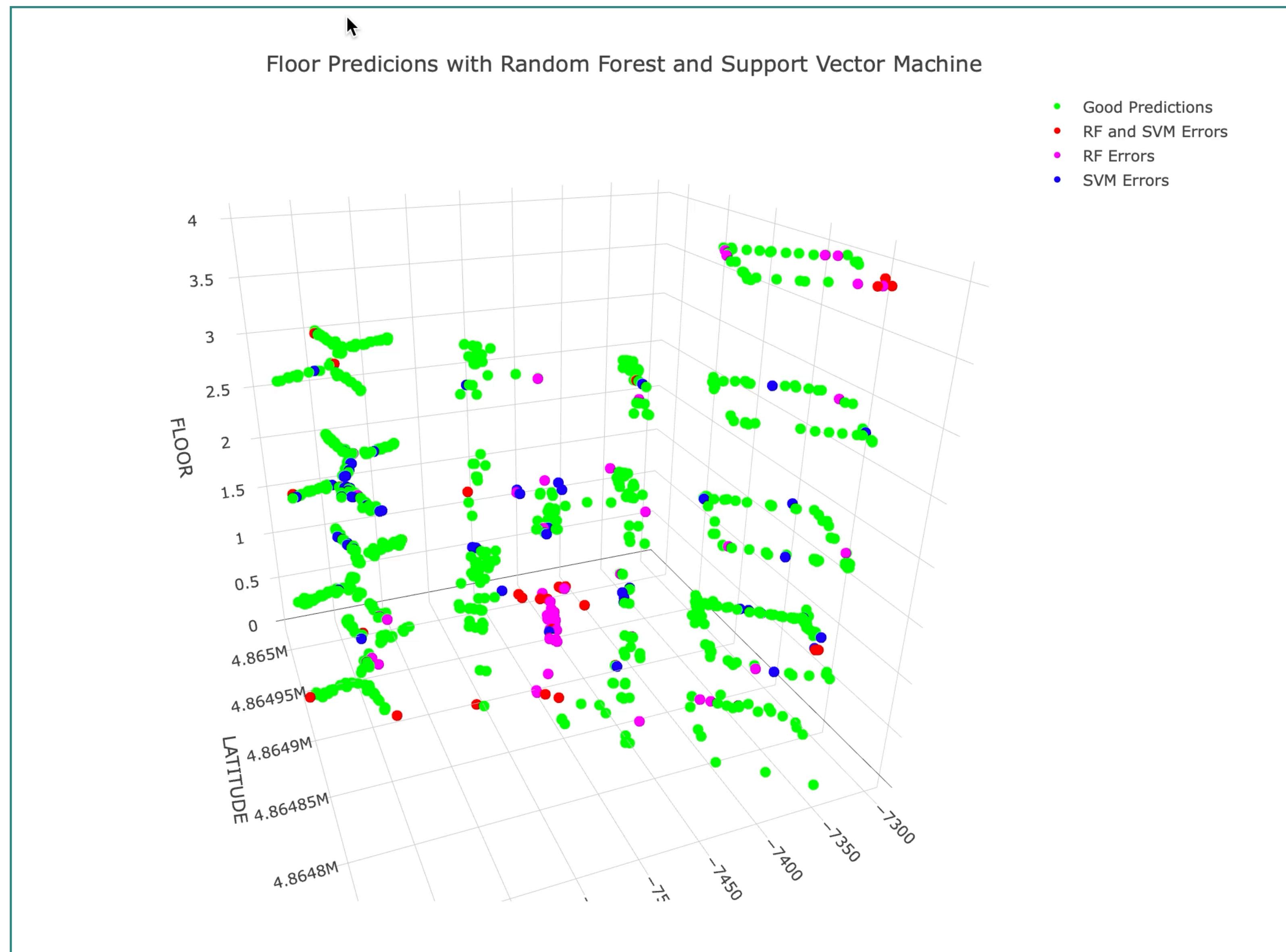
FLOOR PREDICTIONS

Floor		Confusion Matrix						
Support Vector Machine [svm]	Kernel: Linear		SVM					
			Prediction	0	1	2	3	4
	Accuracy	0,911	0	123	24	1	0	0
	Kappa	0,876	1	8	421	28	0	0
Random Forest [RF]	ntree: 100, mtry: 44		2	1	15	269	7	0
	Accuracy	0,914	3	0	2	8	164	3
	Kappa	0,880	4	0	0	0	1	36
			RF	Reference				
Random Forest [RF]			Prediction	0	1	2	3	4
	Accuracy	0,914	0	115	3	0	0	1
	Kappa	0,880	1	10	412	4	0	1
			2	6	42	293	5	0
			3	1	5	9	166	7
			4	0	0	0	1	30

FLOOR ERROR ANALYSIS

Building 2 – Floor 1 Building 3 – Floor 4

- Not enough “good” fingerprints for each spot.
- The **RSSI** for both locations are **weak** and the mobile phones are detecting WAP’s in other floors or buildings.
- **Problems** with the **mobiles** devices not **detecting** the **same WAP’s** per location.
- The **shape** and location **of the buildings** affects the detection of WAP’s, specifically in Building 2.



LONGITUDE AND LATITUDE PREDICTIONS

	Longitude		Latitude	
	K-NN	Random Forest	K-NN	Random Forest
Parameters	K = 5	ntree: 100 mtry: 87	K = 5	ntree: 100 mtry: 173
RMSE	11.68	10.87	10.36	10.42
Rsquared	0.99	0.99	0.98	0.98
MAE	6.08	7.15	5.65	6.54

LONGITUDE ERROR ANALYSIS

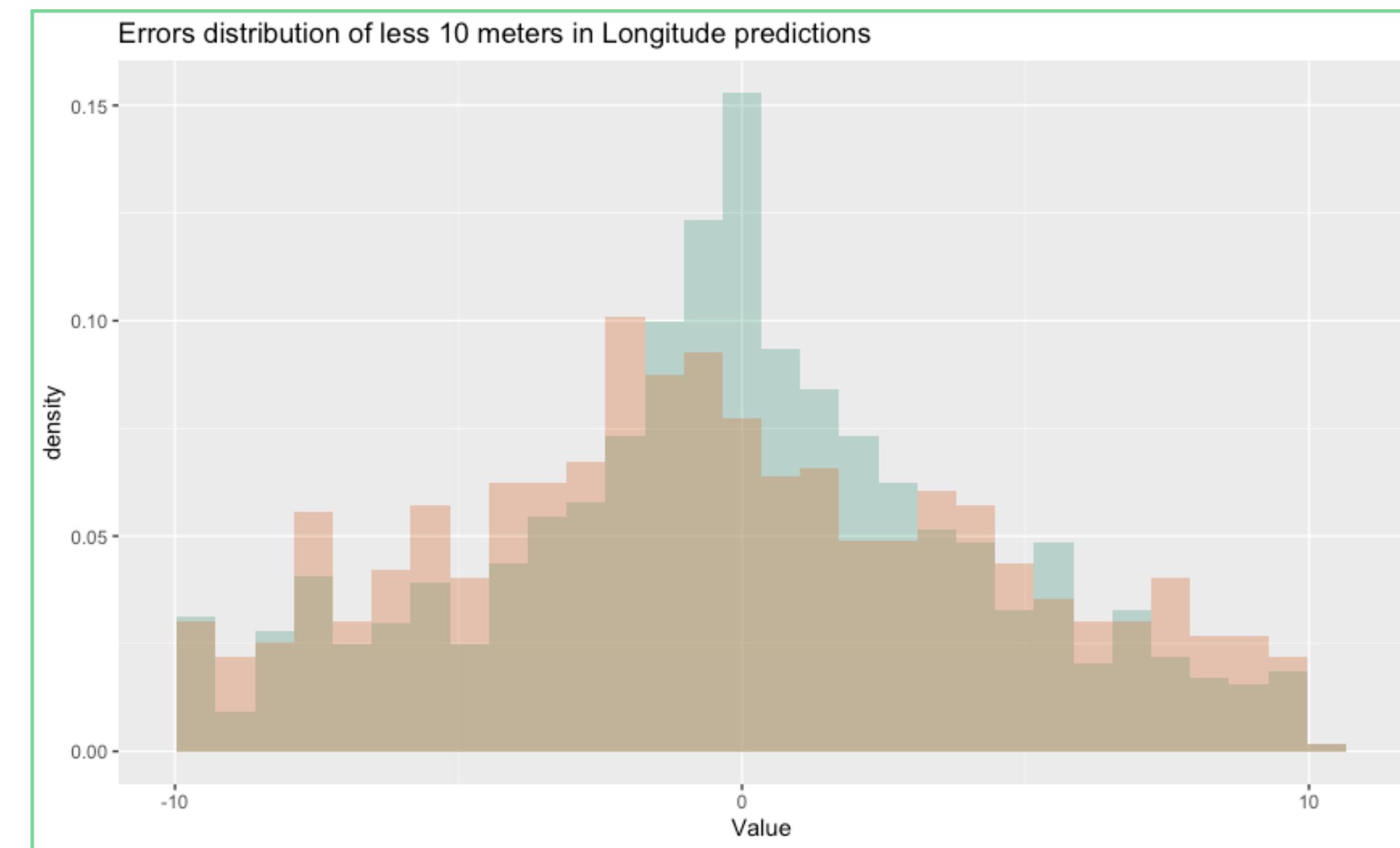
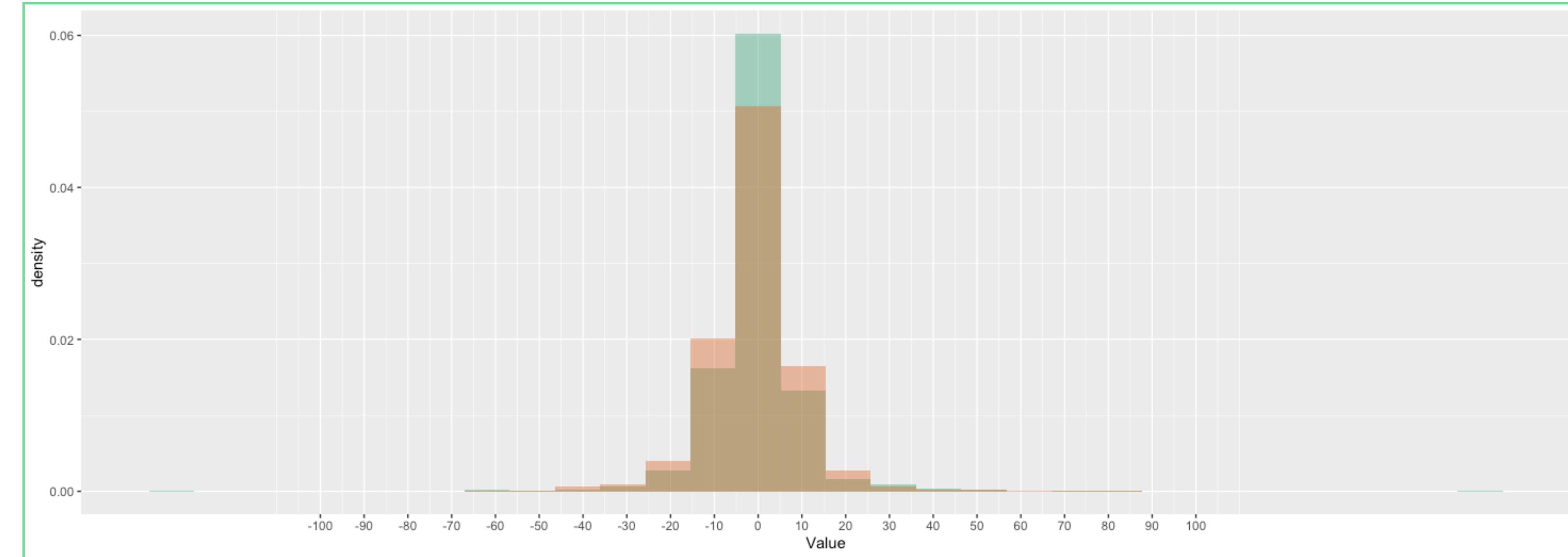
K-NN

Residuals has more density in errors of less than 10 meters and a two outliers in 137 and 164 meters of error.

Random Forest [RF]

Residuals has less density in errors of less than 10 meters and some outliers of 90 meters of error.

	Longitude	
	K-NN	RF
RMSE	11.68	10.87
Rsquared	0.99	0.99
MAE	6.08	7.15
Median	3.45	4.79



K-NN
 Random Forest

LATITUDE ERROR ANALYSIS

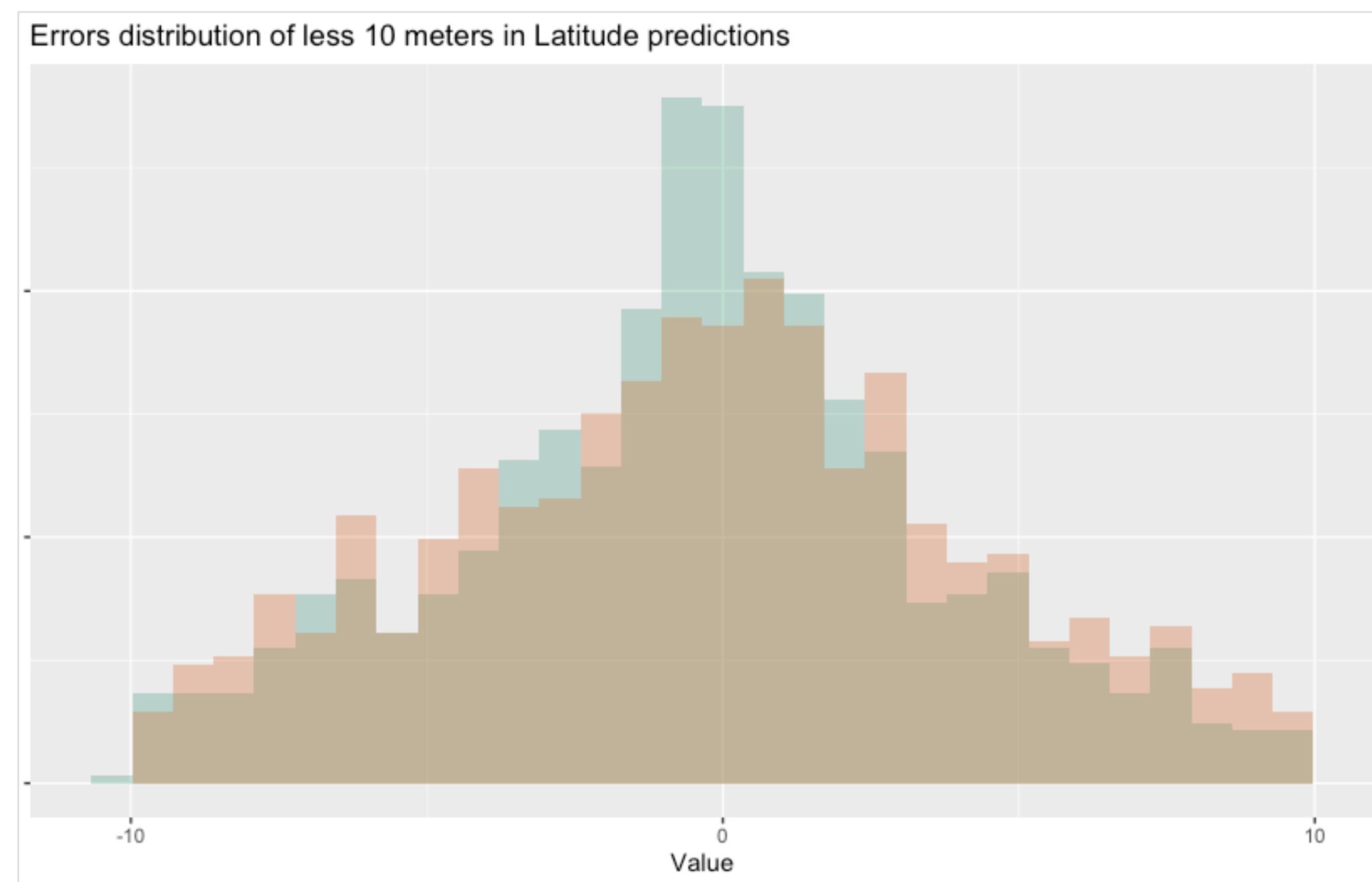
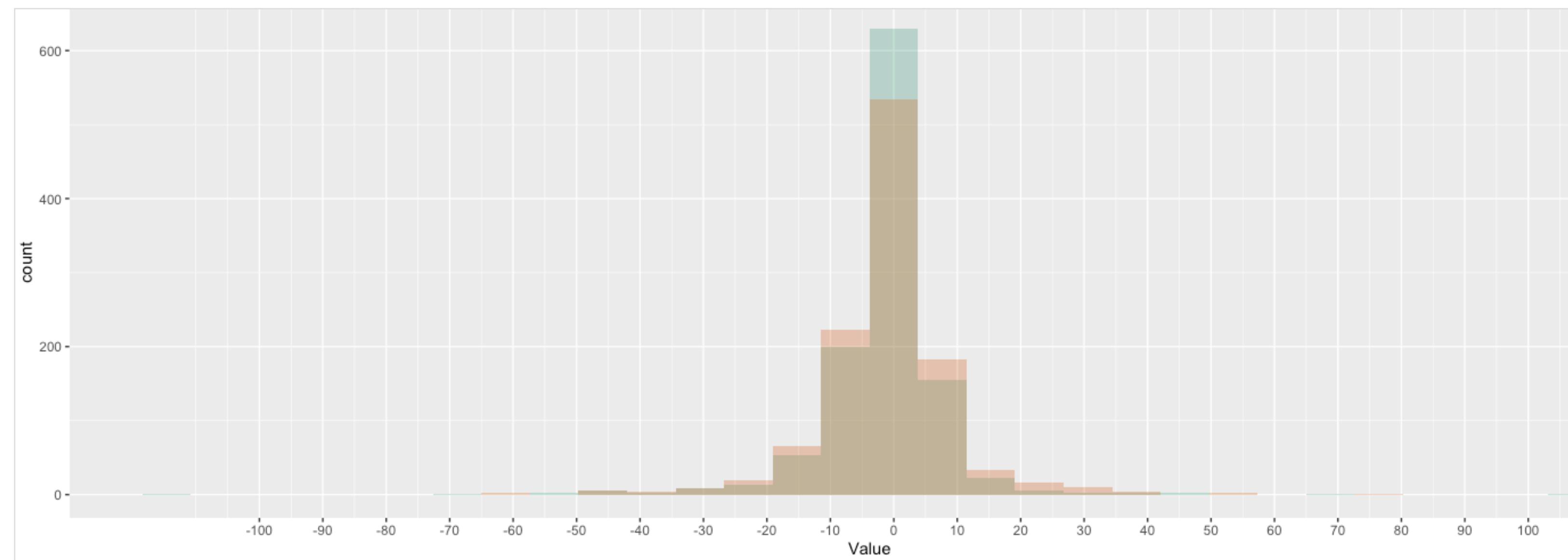
K-NN

Residuals has more density in errors of less than 10 meters and a few outliers in 107 meters of error.

Random Forest [RF]

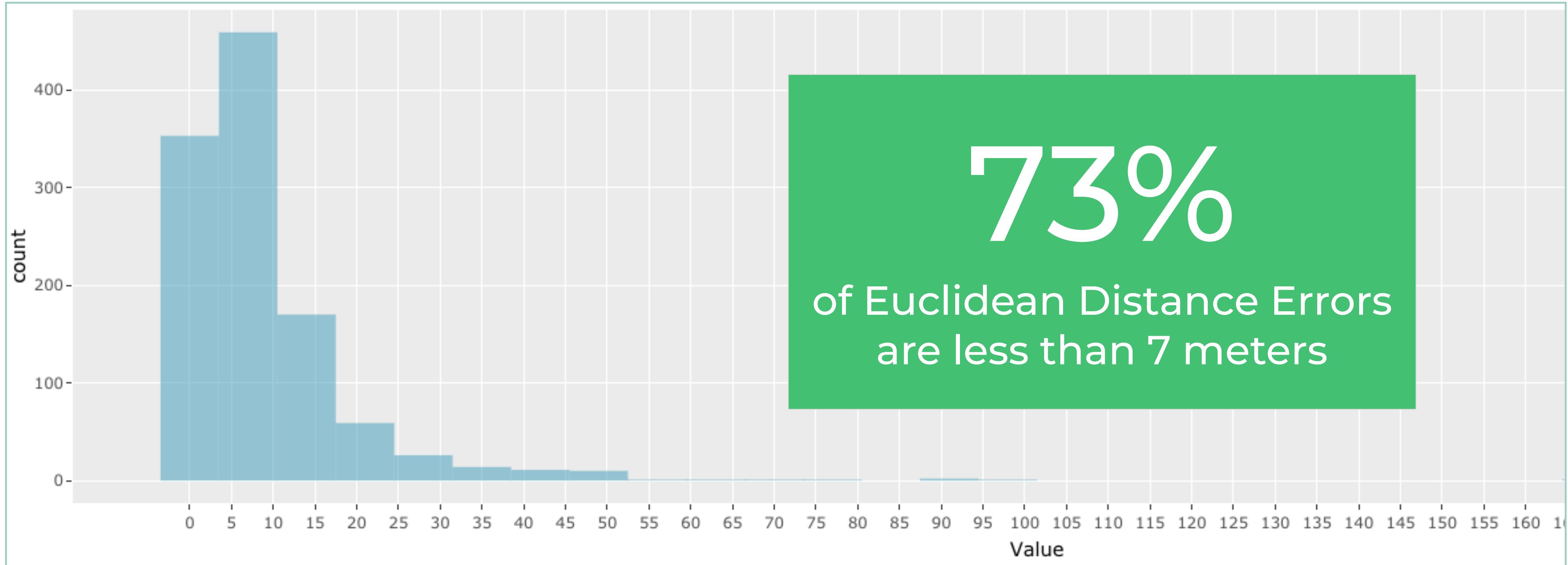
Residuals has less density in errors of less than 10 meters and some outliers of 80 meters of error.

	Latitude	
	K-NN	RF
RMSE	10.36	10.42
R squared	0.98	0.98
MAE	5.65	6.54
Median	3.07	3.99



K-NN
 Random Forest

EUCLIDEAN DISTANCE ERRORS



RESULTS

Longitud

Mean distance error

6,1 meters

Median distance error

3,5 meters

Latitud

Mean distance error

5,6 meters

Median distance error

3,1 meters

Euclidean distance

Mean distance error

9,1 meters

Median distance error

5,9 meters

RECOMMENDATIONS

- Analyze and **Identify** relocated WAPS.
- **Combine** wifi with **Bluetooth** technology.
- **Analyze** each **cellphone** and the signals received by device; cloud be some issues with the samples taken by some phones.
- Perform a **PCA analysis**, this will reduce the dimensionality of our dataset and improve the computational time when applying the algorithms.
- Perform an **analysis** using **H2O**, which is a powerfull algorithm to work with large datasets

T H A N K S !