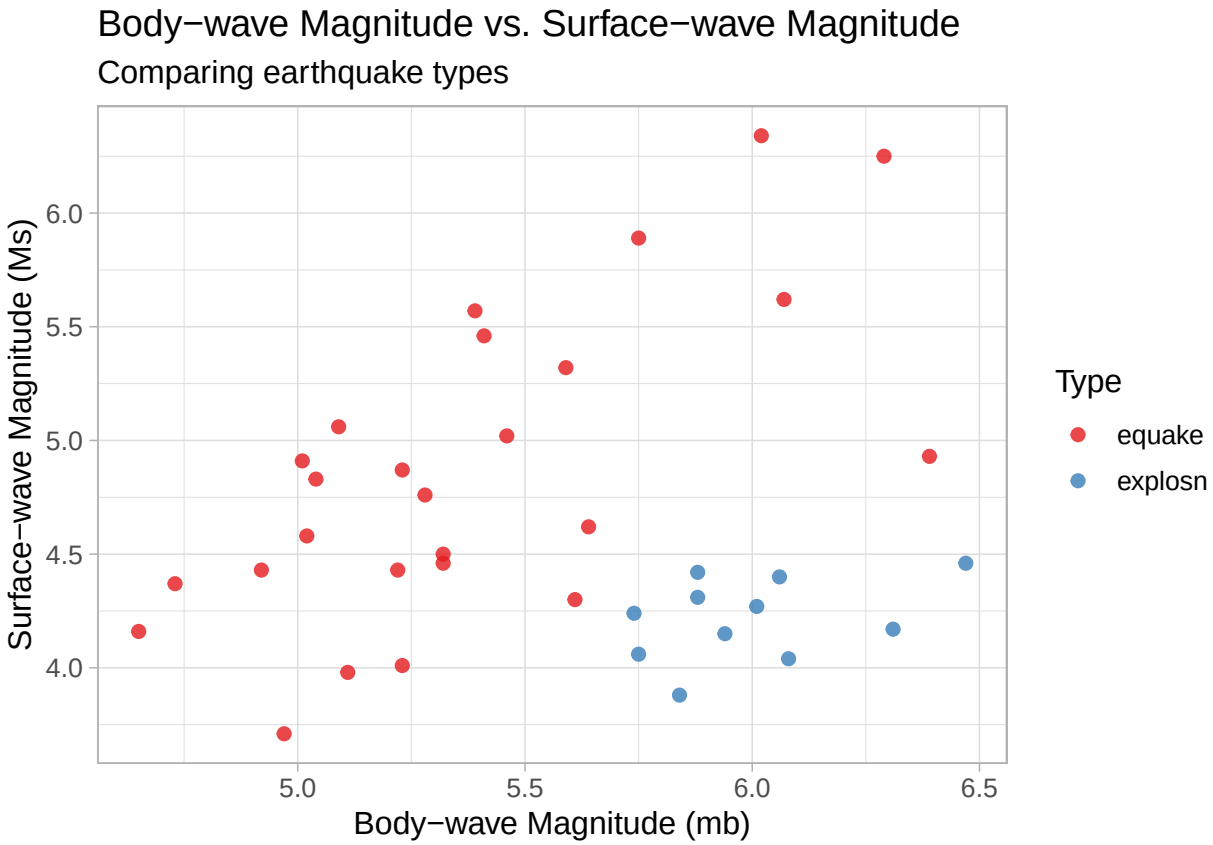
Report_10883408

10883408

2024-04-08

ML part A



Graphical Analysis:

The scatter plot created by this code would show each earthquake or explosion event as a point in the space defined by its body-wave and surface-wave magnitudes. The colors distinguish between different types of seismic events, which could be crucial for identifying patterns or clusters specific to earthquakes versus explosions.

Clustering: If there are visible clusters or distinct areas predominantly occupied by one type of event, this could indicate that these two features (body-wave and surface-wave magnitudes) are effective for distinguishing between earthquakes and explosions.

Overlaps: Significant overlapping of colors might suggest that the two features alone are not sufficient to distinguish between the event types without additional information or more complex modeling.

Contextual Relevance:

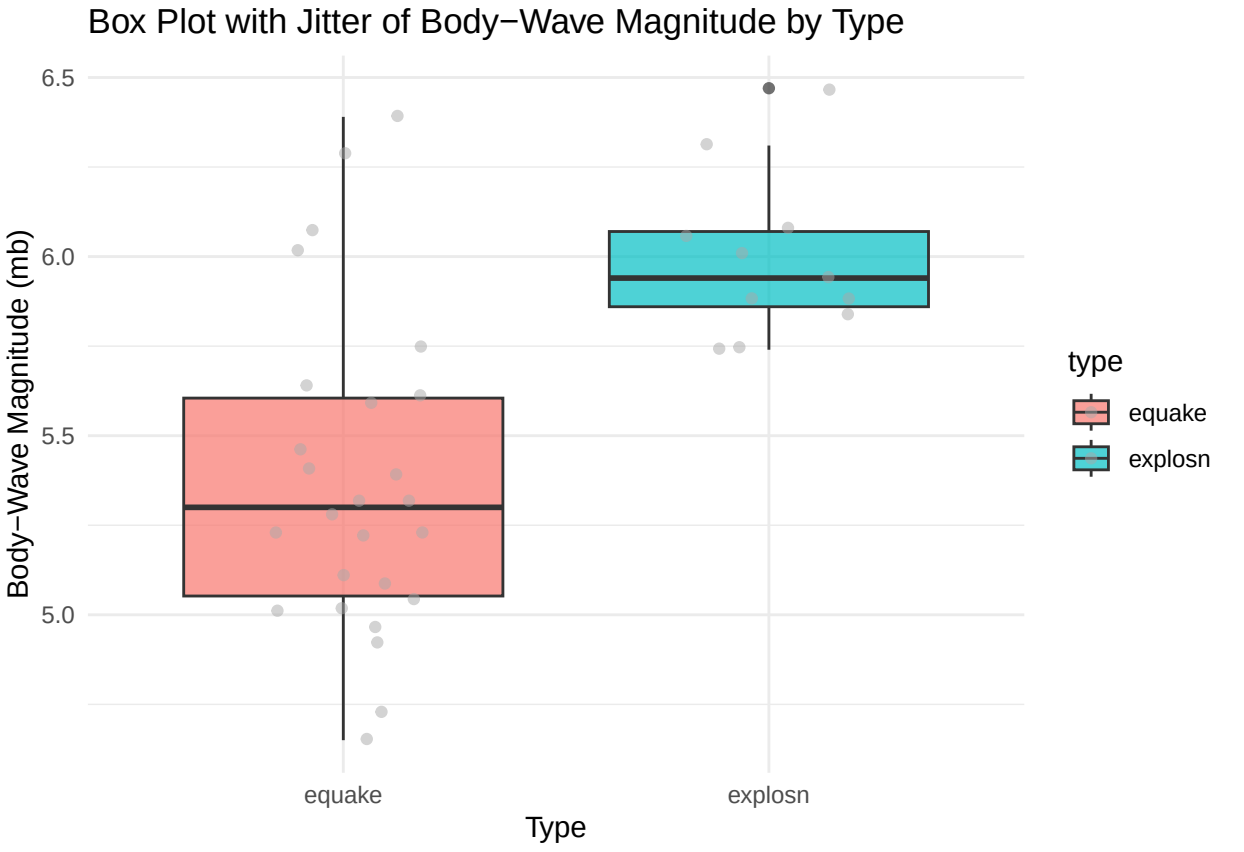
In the context of monitoring for unauthorized nuclear tests, this visualization helps in quickly assessing whether there are clear, distinguishable patterns in seismic readings that could indicate nuclear activities. Effective differentiation between natural seismic events (earthquakes) and man-made seismic events (nuclear explosions) is crucial for global security and monitoring compliance with international treaties such as the Comprehensive Nuclear-Test-Ban Treaty (CTBT).

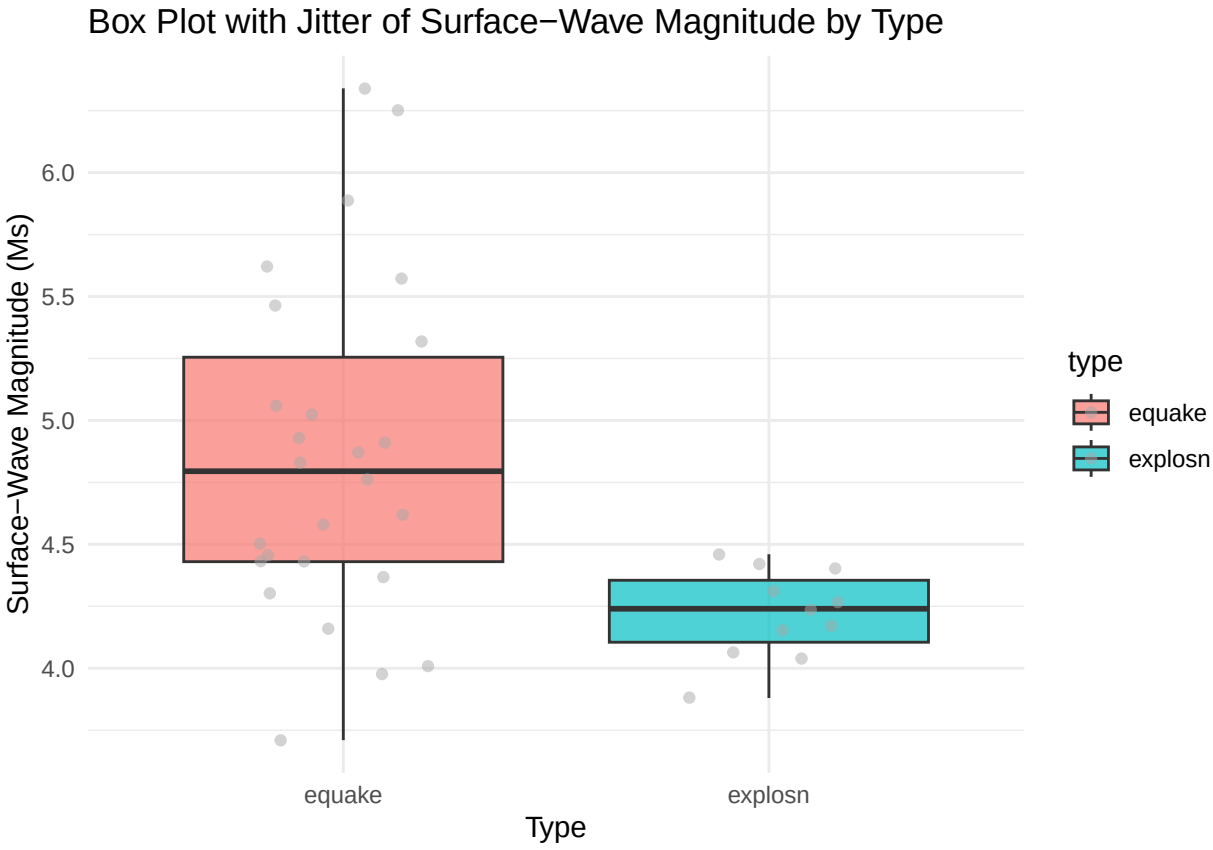
Numerical Summaries:

While the provided code focuses on visual analysis, numerical summaries (like the mean, median, variance of mb and Ms for each type) would complement this by quantifying the central tendencies and dispersions. This could further aid in understanding how significantly the magnitudes for each type differ statistically.

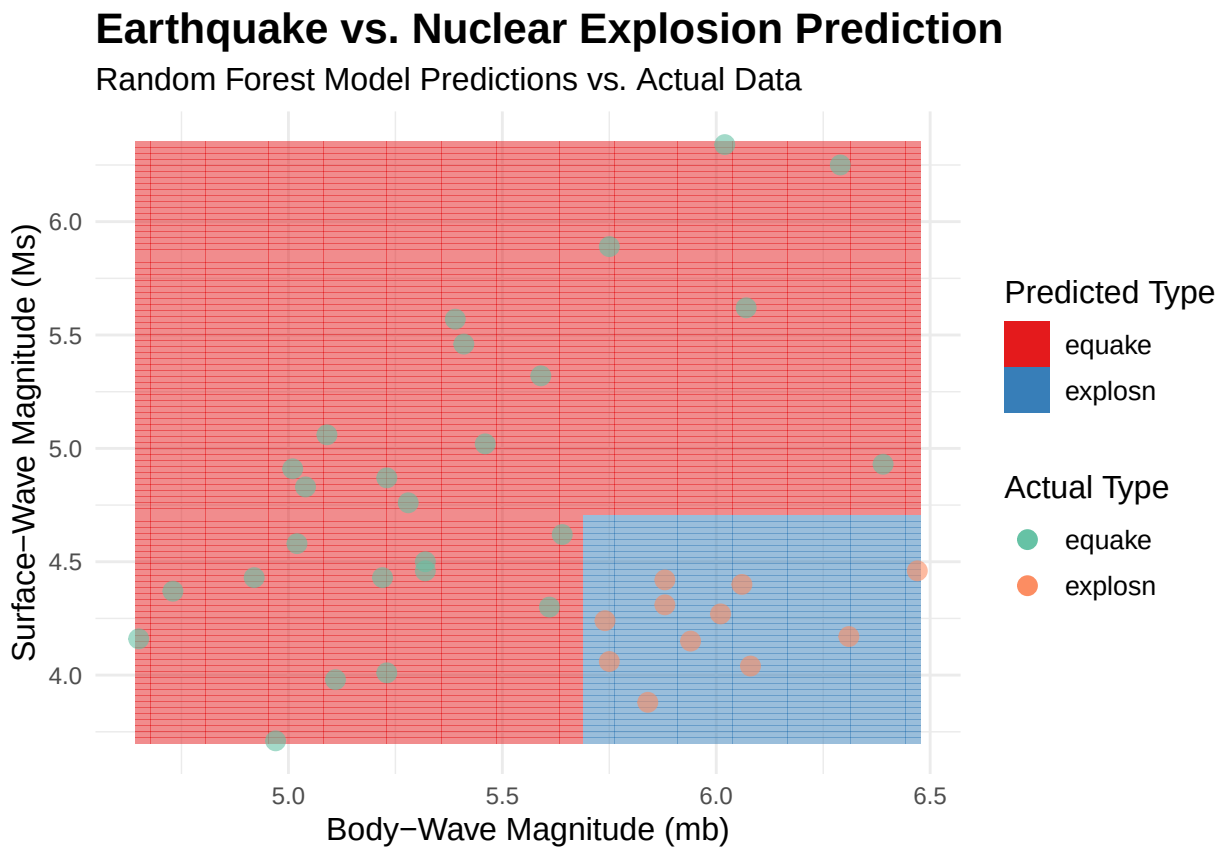
Justification:

The choice of a scatter plot is justified as it allows stakeholders to visually parse the relationship between two continuous variables across categories. Given the high stakes involved in nuclear monitoring, quick visual assessments alongside rigorous statistical analysis are imperative. The plot facilitates this by providing a clear, immediate visual summary of the data as per the described features.



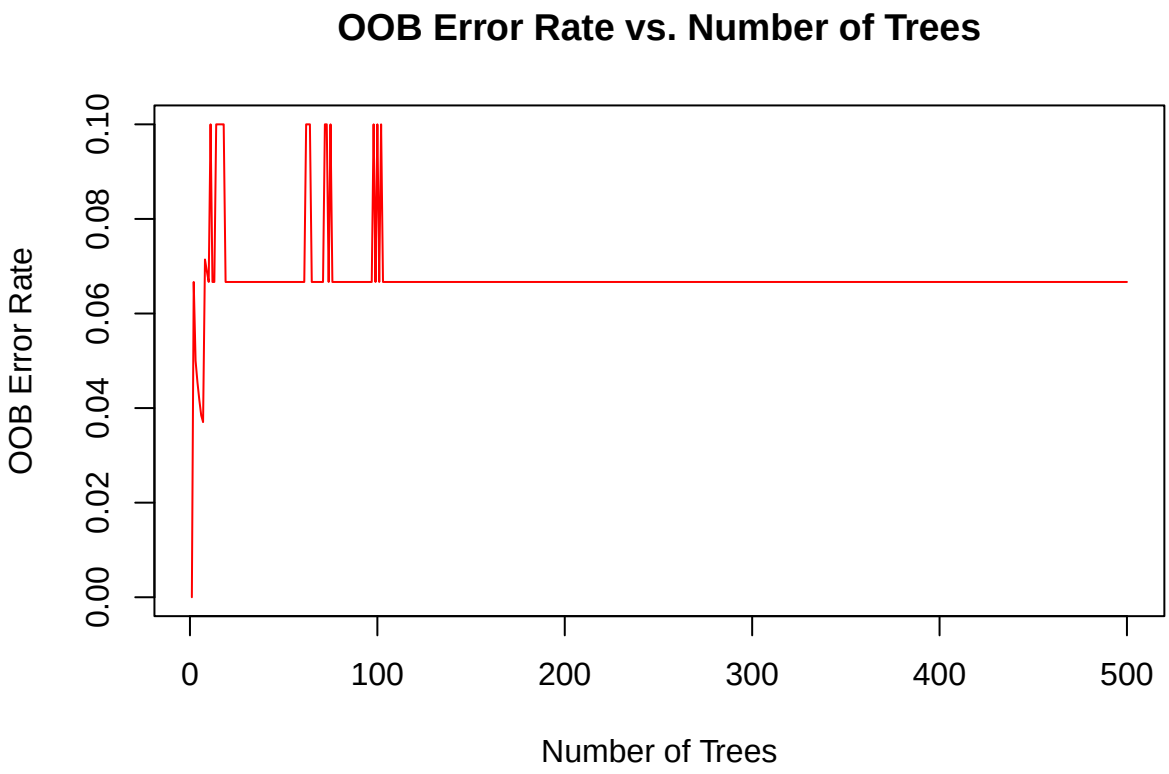

Box Plot with Jitter of Surface−Wave Magnitude by Type
Surface−Wave Magnitude (Ms)
6.0
5.5
5.0
4.5
4.0
equake
explosn
Type
type
equake
explosn

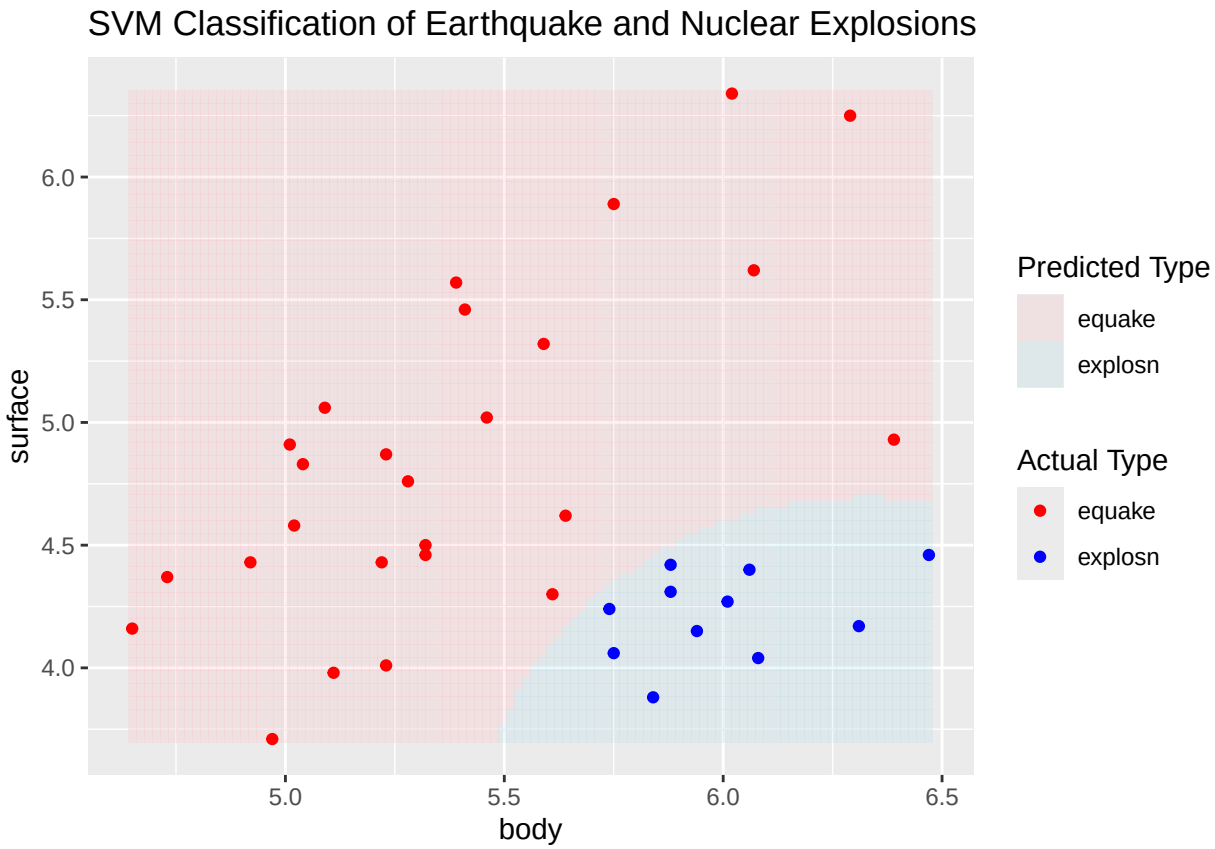
ML part B



error rate

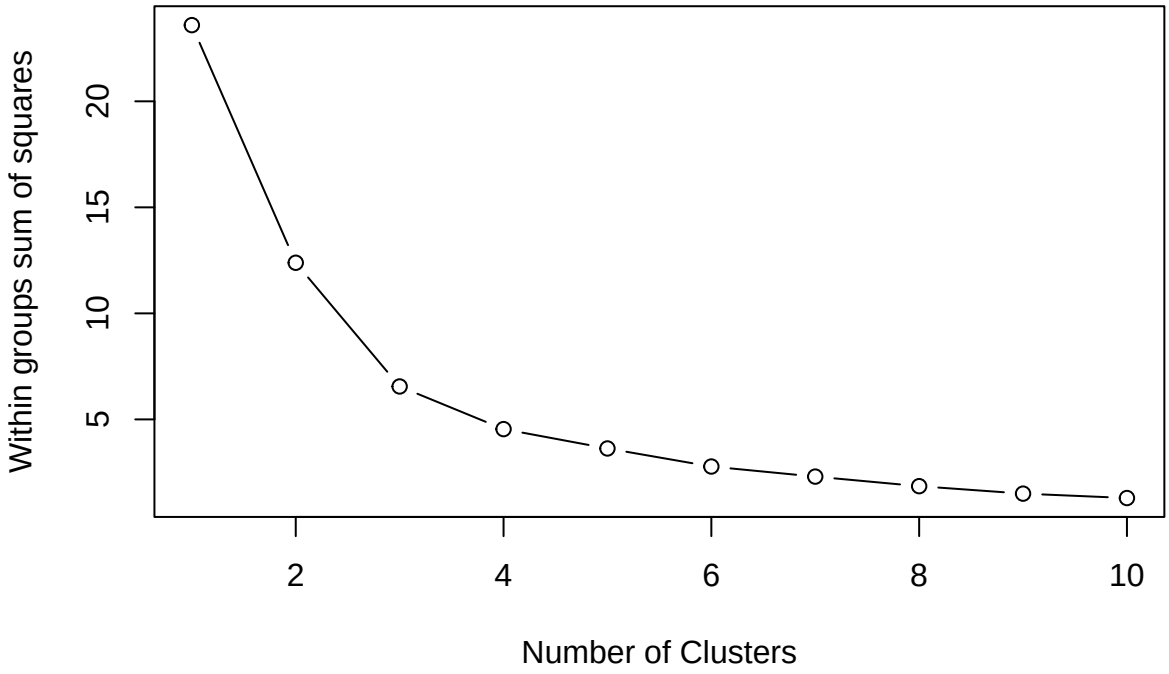
```
# Extract the Random Forest model from the caret object
rf <- rfModel$finalModel

# Error rate plot
plot(rf$err.rate[, "OOB"], type = "l", col = "red",
     xlab = "Number of Trees",
     ylab = "OOB Error Rate",
     main = "OOB Error Rate vs. Number of Trees")
```

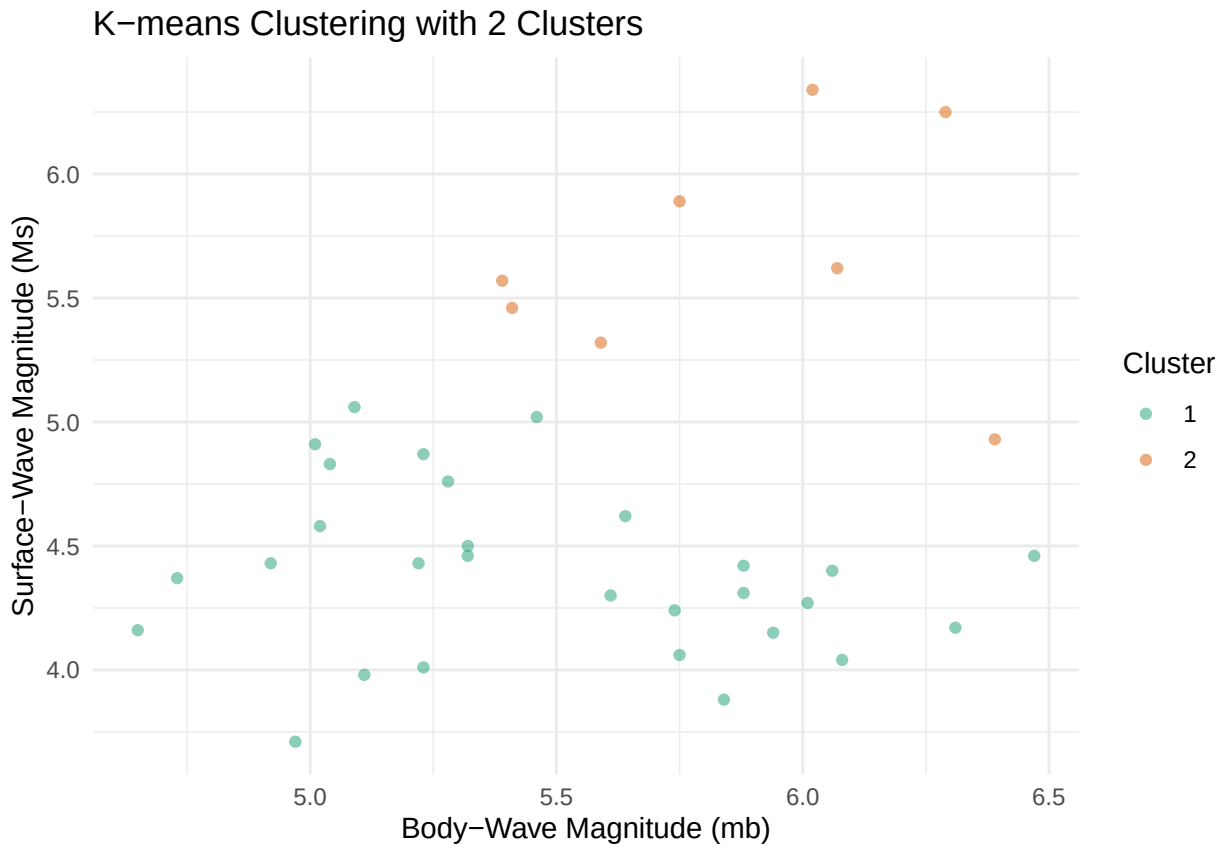

OOB Error Rate vs. Number of Trees
OOB Error Rate
0.00
0.02
0.04
0.06
0.08
0.10
0
100
200
300
400
500
Number of Trees

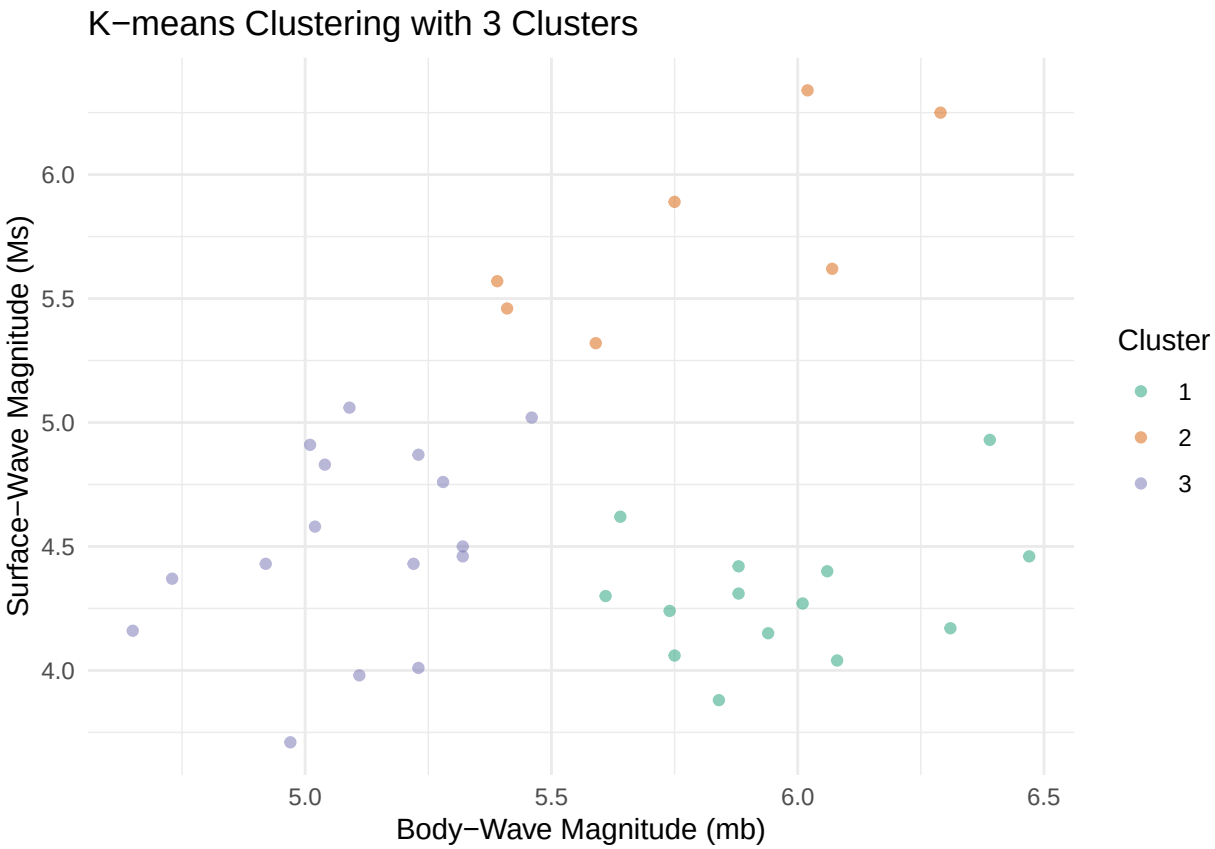


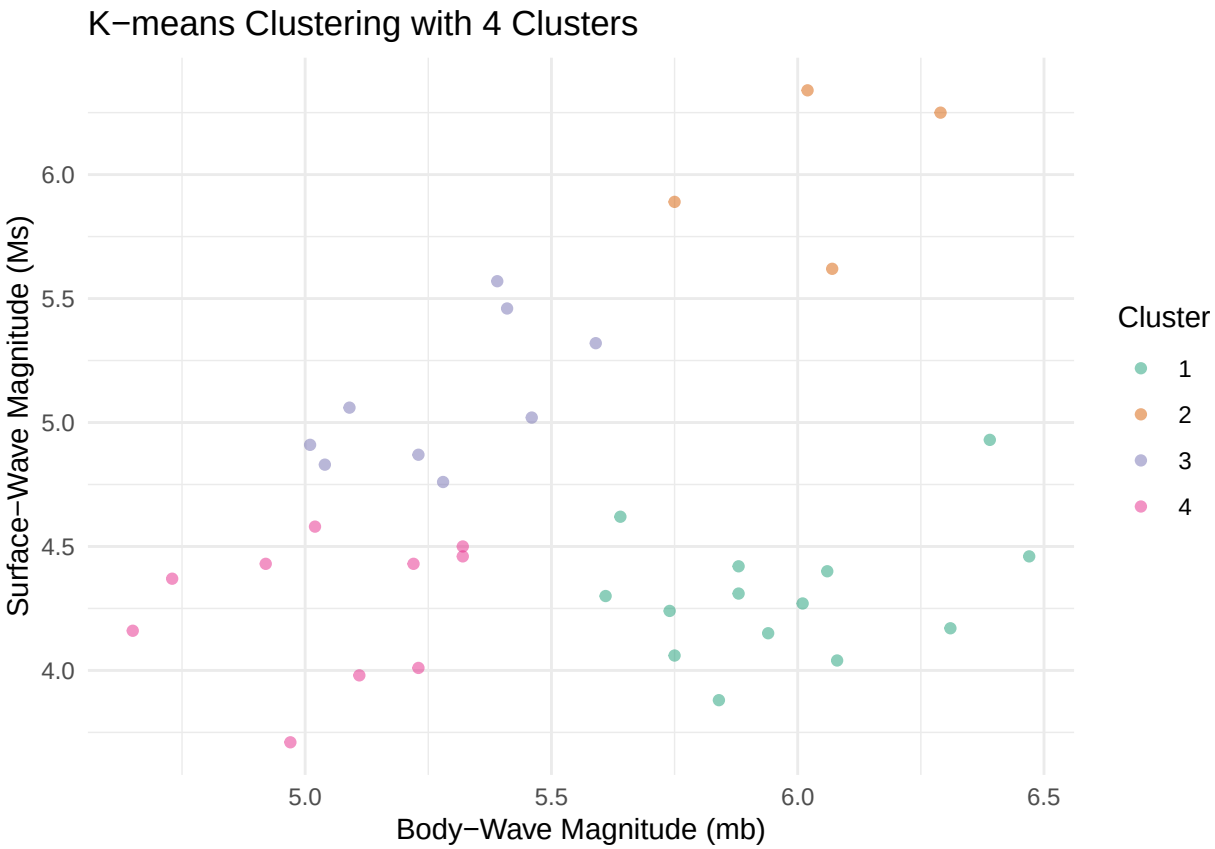

```
##   sigma   C
## 8   0.1 100
```

Part D


Within groups sum of squares
5
10
15
20
2
4
6
8
10
Number of Clusters


K−means Clustering with 2 Clusters
Surface−Wave Magnitude (Ms)
6.0
5.5
5.0
4.5
4.0
5.0
5.5
6.0
6.5
Body−Wave Magnitude (mb)
Cluster
1
2


K−means Clustering with 3 Clusters
Surface−Wave Magnitude (Ms)
6.0
5.5
5.0
4.5
4.0
5.0
5.5
6.0
6.5
Body−Wave Magnitude (mb)
Cluster
1
2
3


K−means Clustering with 4 Clusters
Surface−Wave Magnitude (Ms)
6.0
5.5
5.0
4.5
4.0
5.0
5.5
6.0
6.5
Body−Wave Magnitude (mb)
Cluster
1
2
3
4