

Visualizing Pangolin Uncertainty (a.k.a. Pangoluncertainty)

Devan Becker

2021-02-10

```
library(here)
library(ggplot2)
library(dplyr)
```

Goal

Visualize/summarize the results of sending resampled uncertainty matrices through pangolin.

Vis Setup

Each row in the dataset represents a single sample from one uncertainty matrix (only dealing with one accession number for now). The columns are as follows:

- **taxon:** the accession name, followed by “.0” for the consensus sequence and “.1” to “.1000” for the 1,000 samples from the matrix.
- **lineage:** The called lineage from pangolin
- **probability:** the bootstrap support for this lineage call
- **pangoLEARN_version, status, and note:** extra info from pangolin

```
unc1 <- read.csv(here("data/pangolineages/", "ERR4085809_pangolineages.csv"))
head(unc1)
```

##	taxon	lineage	probability	pangoLEARN_version	status	note
## 1	_ERR4085809.0	B.1	1	2021-02-21	passed_qc	NA
## 2	_ERR4085809.1	A.2	1	2021-02-21	passed_qc	NA
## 3	_ERR4085809.2	B	1	2021-02-21	passed_qc	NA
## 4	_ERR4085809.3	B	1	2021-02-21	passed_qc	NA
## 5	_ERR4085809.4	B.1.98	1	2021-02-21	passed_qc	NA
## 6	_ERR4085809.5	B	1	2021-02-21	passed_qc	NA

To prep the data, I calculate another column called ""