

Shigella Draft Analysis

1 Matching Compiled Shigella Dataset and 'Actual Day' Dataset

1.1 Checking Compiled Shigella Dataset

```
[1] 48
```

```
# A tibble: 218 x 2
  sid      timepoint
  <chr>      <dbl>
1 SOSAR-11001      2
2 SOSAR-11001      7
3 SOSAR-11001     30
4 SOSAR-11001     90
5 SOSAR-11001    180
6 SOSAR-11002      2
7 SOSAR-11002      7
8 SOSAR-11002     30
9 SOSAR-11002     90
10 SOSAR-11006      2
# i 208 more rows
```

The compiled Shigella dataset contains 48 unique sample IDs with a total of 218 observations.

1.2 Checking New 'Actual Day' Dataset

```
[1] 139
```

The new 'Actual Day' dataset contains 139 unique sample IDs with a total of 428 observations.

1.3 Subsetting the New Dataset Based on the Compiled Shigella Dataset's Sample IDs

```
# A tibble: 229 x 2
  CaseID      `Actual day`
  <chr>         <dbl>
1 SOSAR-11001         5
2 SOSAR-11001        29
3 SOSAR-11001        96
4 SOSAR-11001       202
5 SOSAR-11001       267
6 SOSAR-11001       363
7 SOSAR-11002         5
8 SOSAR-11002        24
9 SOSAR-11002        90
10 SOSAR-11002       206
# i 219 more rows
```

Since the new dataset has more unique sample IDs than the compiled Shigella dataset, we filtered the new dataset to match the sample IDs in the compiled dataset. This allows for a meaningful comparison between the two datasets under the same conditions.

After filtering, the new dataset has 48 unique sample IDs with a total of 229 observations, compared to the compiled dataset's 48 unique sample IDs with 218 observations.

Therefore, we identified an 11-observation difference between the two datasets.