

UK lineages summary report

Traceback (most recent call last):

```
File "/Users/s1743989/anaconda3/envs/report/lib/python3.7/site-packages/IPython/core/interactiveshell.py", line 3331, in run_code
    exec(code_obj, self.user_global_ns, self.user_ns)
File "<ipython-input-1-a08674dfc154>", line 24, in <module>
    import UK_full_report.utils.basic_descriptions as descrip
File "/Users/s1743989/anaconda3/envs/report/lib/python3.7/site-packages/UK_full_report/utils/__init__.py", line 5, in <module>
    from UK_full_report.utils import *
File "/Users/s1743989/anaconda3/envs/report/lib/python3.7/site-packages/UK_full_report/utils/mapping.py", line 121
    if uk_country == "Wales"
    ^
```

SyntaxError: invalid syntax

```
-----NameError Traceback (most recent call last) in
--> 1 current_date = time.make_current_week(week) 2 3 intro_bigs, intro_smalls, intro_all, count, intro_countries, intro_object_dict, omitted, taxa, new_lineages, taxon_dictionary, most_recent_sample, intro_int_list, unclear_taxa = parse.make_objects(metadata_file) 4 singletons_count, smalls_count = descrip.get_preliminary_info(intro_countries) 5 NameError: name 'time' is not defined
```

Note: the size of a lineage may be due to a low amount of transmission of this lineage, but it is likely also that it just hasn't been sampled as frequently, especially for newer lineages.

Sequences which were replicates or too error-prone were removed from this analysis.

```
-----NameError Traceback (most recent call last) in --> 1
print(str(smalls_count) + " are lineages which only contained five sequences or fewer, and so have been left out of visualisation in the interests of clarity") NameError: name 'smalls_count' is not defined
```

```
-----NameError Traceback (most recent call last) in 1
#split = lin_exp.find_splits(intro_all) 2 --> 3 status_counts, reactivated_lineages, continuing_lineages = lin_exp.describe_lineages(intro_bigs) 4 5 reactivateds = status_counts["Reactivated"] NameError: name 'lin_exp' is not defined
```

The following table contains information about lineages and the number of sequences in each country in the UK for each lineage, in reverse size order. Each entry is the count of sequences from each lineage in each country, with the percentage of the total sequences from that lineage that this count represents.

It is also written to "summary_files" as "lineage_summary.tsv" for further use.

```
-----NameError Traceback (most recent call last) in
--> 1 intro_country_counts, intro_country_percentages, intro_country_together, intro_total_numbers = descrip.prep_dicts(intro_countries) 2 dataframe, tree_order = descrip.make_dataframe(intro_country_together, intro_total_numbers, intro_object_dict) 3 4 print(dataframe.to_markdown()) NameError: name 'descrip' is not defined
```

These data is represented in the stacked bar chart below. Note that the number of sequences is likely to be due more to differing sampling efforts in different regions, rather than genuine differences in numbers of cases.

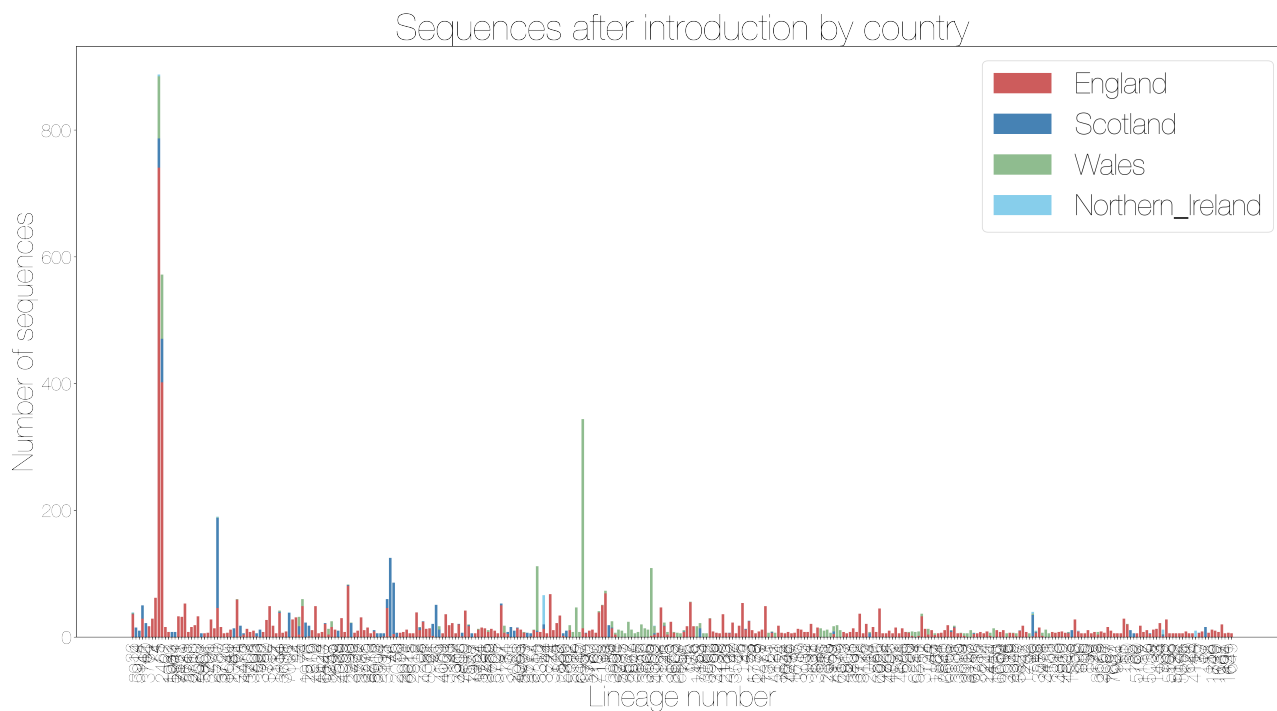
The raw data for this bar chart are in the table above.

```
-----NameError
```

```

Traceback (most recent call last)<ipython-input-1-82848448dadab> in
<module>
----> 1 df_counts, df_thinned, df_acctrans_counts =
dp.make_plotting_dfs(intro_country_counts, intro_object_dict)
      2
      3 dp.plot_bars(intro_bigs)
NameError: name 'dp' is not defined

```

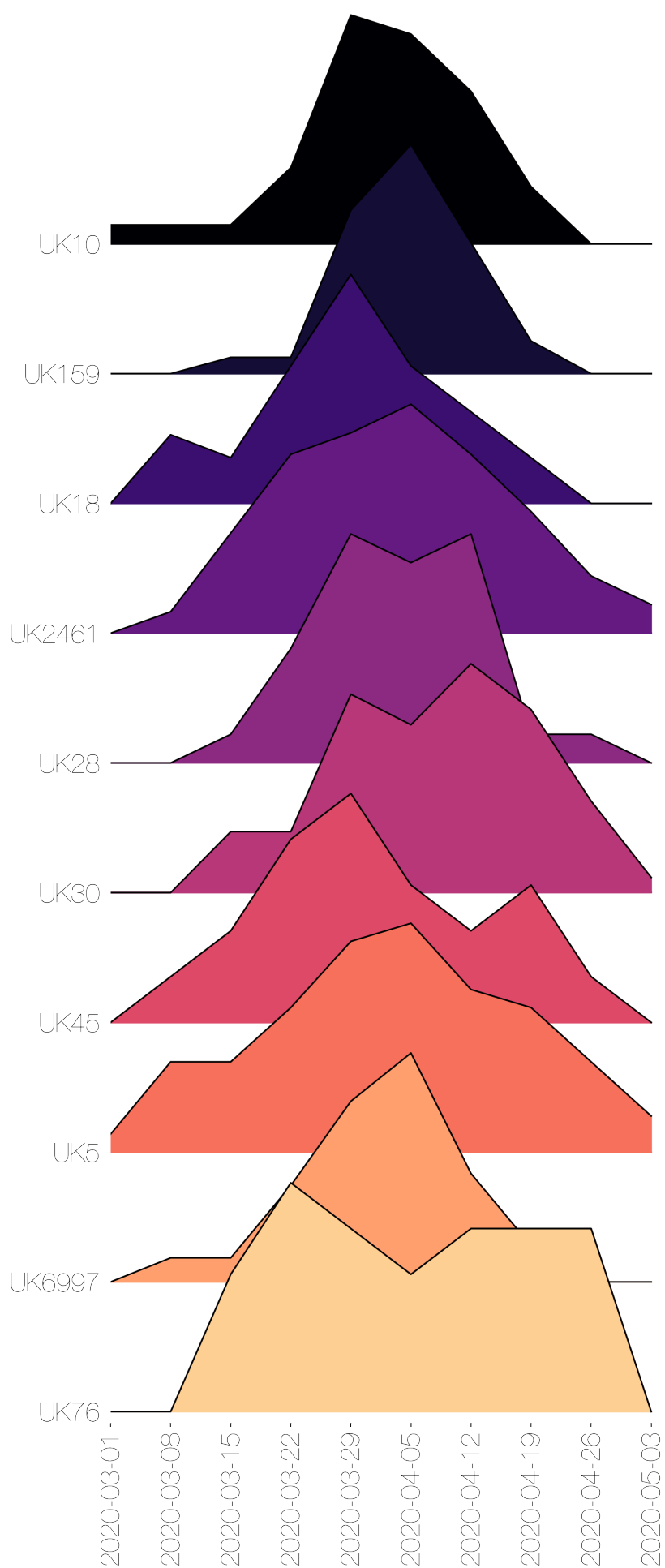


The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown below. The raw data for the plot is shown below it, with each column representing a lineage, and the number of admin2 regions it is present in in each week.

```

-----NameError
Traceback (most recent call last)<ipython-input-1-865e61bd73ac> in
<module>
----> 1 y_dict, week_list = dp.prep_geog_data(intro_bigs)
      2 #dp.stacked_geog_plot(y_dict, week_list, False)
      3 dp.plot_ridge_plot(week_list, y_dict)
      4 raw_geog = dp.make_raw_data_geog_plot(y_dict, week_list)
NameError: name 'dp' is not defined

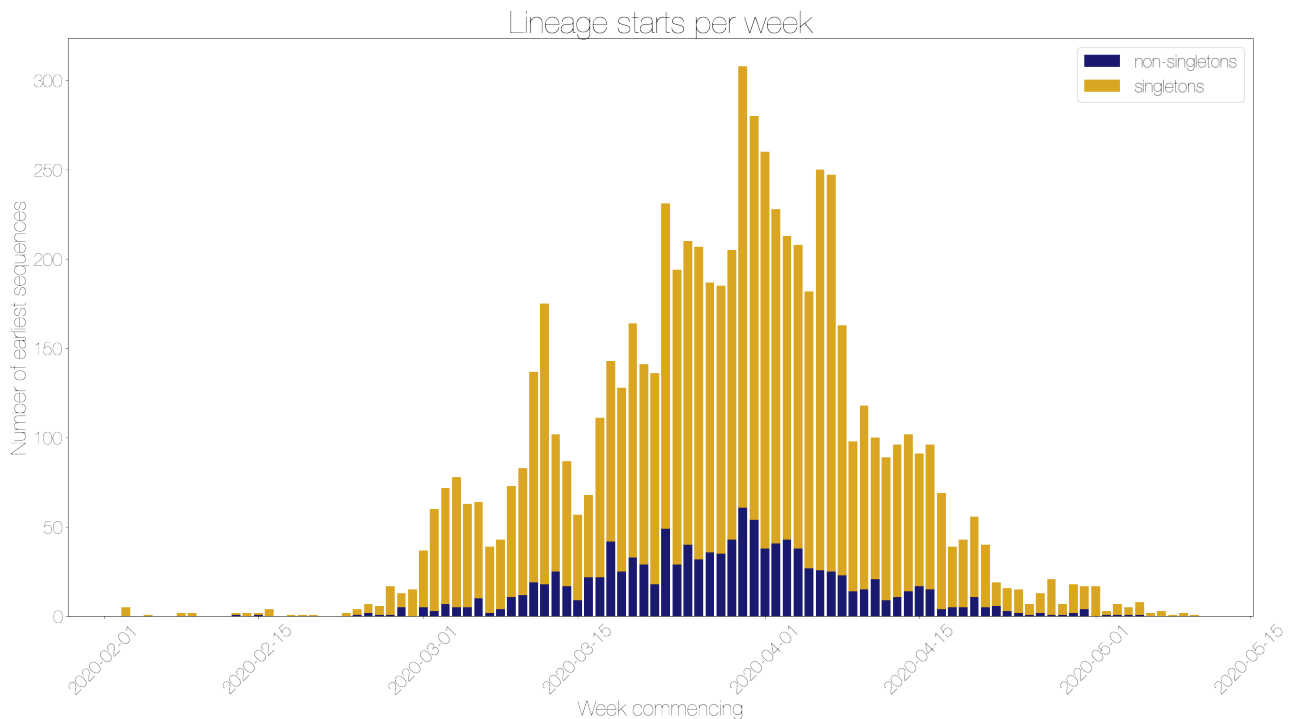
```



```
-----NameError Traceback (most recent call last) in 1
print("{}( + fd + "/" + name_stem + "_geog_plot_1.png)\") --> 2 print(raw_geog.to_markdown()) NameError:
name 'raw_geog' is not defined
```

The date of first sequence in the cluster is shown below for every cluster with date information.

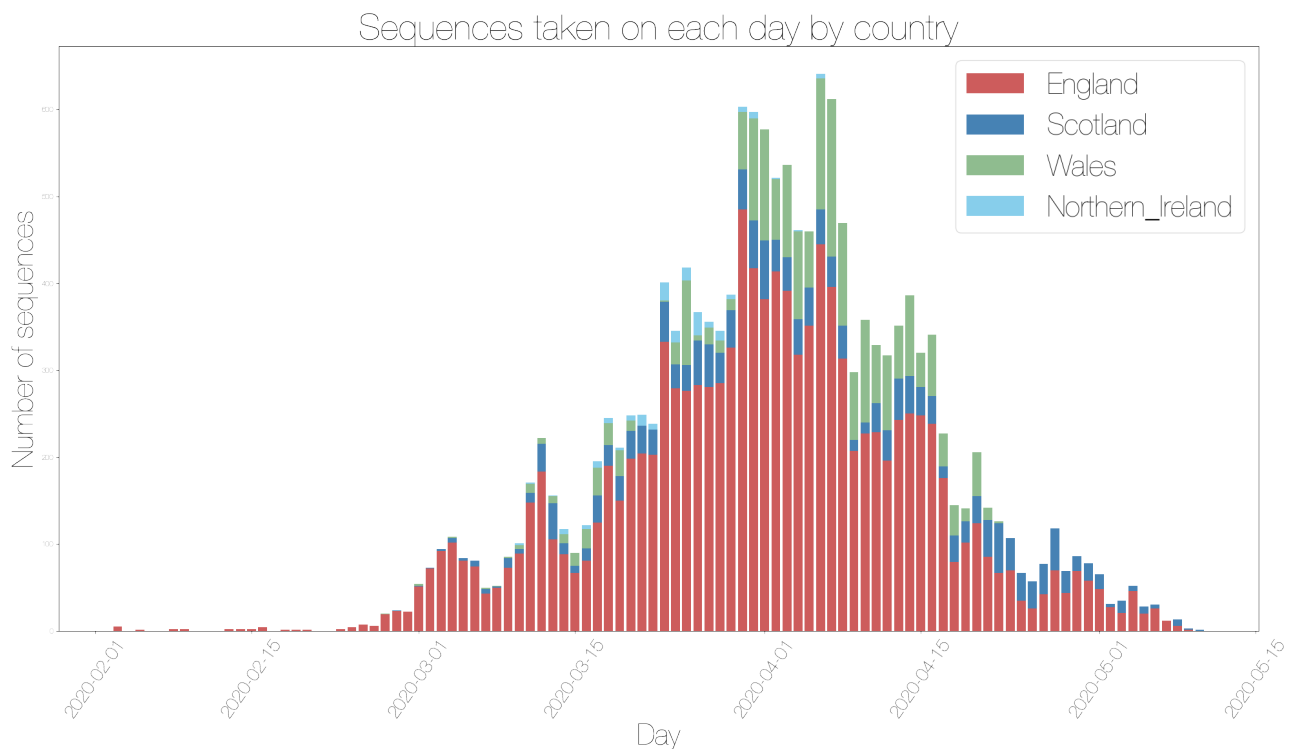
```
-----NameError
Traceback (most recent call last)<ipython-input-1-5f5882fbc34d> in
<module>
----> 1 multi, single = dp.plot_starts(intro_all)
      2 starts_raw = dp.raw_data_starts(single, multi)
NameError: name 'dp' is not defined
```



```
-----NameError Traceback (most recent call last) in 1
print("{}( + fd + "/" + name_stem + "_firsts_plot_1.png)\") --> 2 print(starts_raw.to_markdown()) NameError:
name 'starts_raw' is not defined
```

For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.

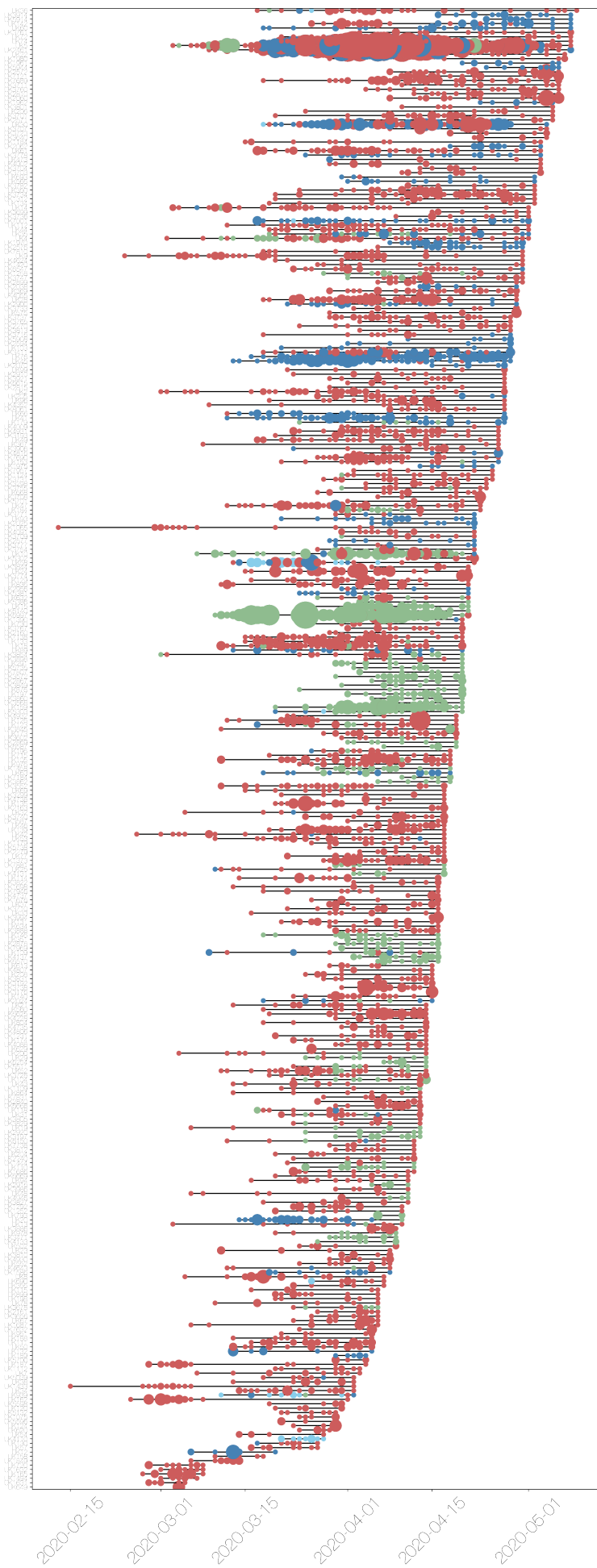
```
-----NameError
Traceback (most recent call last)<ipython-input-1-1bbf7b658203> in
<module>
----> 1 days, E, S, W, NI = dp.plot_sequences_over_time(taxa)
      2 raw_seqs_over_time = dp.raw_data_seqs_over_time(days, E, S, W,
NI)
NameError: name 'dp' is not defined
```



```
-----NameError Traceback (most recent call last) in 1
print("{}( " + fd + "/" + name_stem + "_seqs_over_time_1.png)") --> 2 print(raw_seqs_over_time.to_markdown())
NameError: name 'raw_seqs_over_time' is not defined
```

These lineages are shown on the timeline below. Each line represents the length of the cluster, from oldest to most recent sampling date. The dots are sized by the number of sequences taken on that date, and again are colour coded by country. The raw data has been written to a summary file.

```
-----NameError Traceback (most recent call last) in --> 1
dp.make_timeline(intro_bigs) 2 timeline_df = dp.raw_data_timeline(intro_bigs) NameError: name 'dp' is not
defined
```



The map below shows the number of sequences sampled in each admin2 region in the UK. The colour scale is the same for all four countries, but with different underlying base colours.

```
-----AttributeError Traceback (most recent call last) in 5 in-
put_geojsons = [uk_json, channels, NI_json] 6 --> 7 new_uncleans, mapping_data = map.make_map(input_geojsons,
adm2_cleaning_file, metadata_file, output_directory, week) AttributeError: type object 'map' has no attribute
'make_map'
```

```
-----NameError Traceback (most recent call last) in 1
print("{}( " + fd + "/" + name_stem + "_map_1.png)\") --> 2 print(mapping_data.to_markdown()) NameError:
name 'mapping_data' is not defined
```

```
-----NameError Traceback (most recent call last) in -->
1 if new_uncleans: 2 print("There are some sequences with locations that are not matched to real Admin2
regions, some manual curation required.") NameError: name 'new_uncleans' is not defined
```

Other results modules for UK lineage analysis can be added in here if required.

Appendix

The plot below shows the number of sequences from each country that don't have specific enough location data to plot on the map.

```
-----NameError
Traceback (most recent call last)<ipython-input-1-4758bb6fa688> in
<module>
----> 1 writing.write_summary_files(output_directory, dataframe,
omitted, week, intro_all, timeline_df)
NameError: name 'writing' is not defined
```