Week of Jan 31st 2022

* plot color scheme – form dropdown – so you can view palettes
* upload file section in forms?
* change date formatting so Excel doesn’t convert weeks (force as characters) – uses *YearWeek* instead of just *Week*
* output a table of || TP2 cluster | Pango lineage | Size ||
* check this is easy to extract for user
* want to minimize user text entry
* should be able to get top 10 clusters by size *and* by rate (ones with fastest growth rate)
* clustering before the heatmap
* should be clearly clustered
* pull out the heatmap color palette (should be user-adjustable)
* paired histograms – should be able to shade these
* cluster by (geo, temp, temp-geo) – user choice
* better average\_dists implementation (faster, for larger files)
* go through GZ’s report – fine tune some more – see which parts can be generated by pre-processing scripts, do so
* whisker plots – too much data?
* the whisker plot run out of memory if running all top 20 clusters when there are two cluster sizes bigger than 9000 (comment out)
* heatmaps – original cluster name
* heatmap adjustment – for largest growth clusters (what is minimum size to consider? note: min value of 2, will not consider singletons)
* weekly ECC not consistent with monthly ECC at turn of year
* new report missing the first week of 2021, use *YearWeek* instead of *Week*
* transform in heatmap section should be off as default
  + add as option in form\_inputs
* documents describing the analysis – figure out which components need this (of my part)
* documents describing the analysis – organize how to do this
* update readme
* novel\_only ECC columns not found
* top growth clusters not same as GZ’s
* line 919 – pairwise geo and temp distance for top 20 clusters need to calculate the actual physical pairwise distance, then we have to change the distance calculation in step 4
* modular heatmaps – big task
* should make the heatmap generation as independent of the cgm-ecc-metrics generation as possible