

RSView Technology Review

What do we need to do for our project?

- Download RSV G sequences
- Map RSV sequences based on location collected
 - Color code each point by subtype
 - Change the size of each point based on subtype prevalence
- An interactive map would be a plus
 - Include a slider to adjust the years and/or seasons included on the map
- Overlay additional data on map
 - Also display correlation between subtypes and pneumonia deaths
 - Possibly through alternate color coding options?

Accessing the sequence data from GenBank

NCBI Resources ▾ How To ▾

Nucleotide Nucleotide ▾ "human respiratory syncytial virus" g × Search

[Create alert](#) [Advanced](#)

Items: 1 to 20 of 18370

<< First < Prev Page 1 of 919 Next > Last >>

- ☐ [Human respiratory syncytial virus nonstructural protein \(1C\), nonstructural protein \(1B\), major nucleocapsid \(N\), phosphoprotein \(P\), protein \(M\), 1A \(1A\), G \(G\), protein \(F\) and envelope-associated protein \(22K\) gene, complete cds](#)
- 1.

8,510 bp linear RNA

Accession: M11486.1 GI: 333925

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

- ☐ [Human respiratory syncytial virus \(MON-1-87\) subgroup A, G gene for glycoprotein](#)

2. 922 bp linear RNA

Accession: Z33421.1 GI: 485905

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

Bio.Entrez overview

Pros:

- Designed specifically for downloading data from NCBI
- Bypasses need for URL download
- Customizable downloads
- BioPython and Entrez have good documentation

Cons:

- Object structure not entirely intuitive
- Dictionary and string parsing necessary to get important metadata
- Only option

Human respiratory syncytial virus (MON-1-87) subgroup A, G gene for glycoprotein

GenBank: Z33421.1

[FASTA](#) [Graphics](#)

[Go to:](#) 

```
LOCUS      Z33421                      922 bp    RNA        linear    VRL 15-DEC-1997
DEFINITION Human respiratory syncytial virus (MON-1-87) subgroup A, G gene for
            glycoprotein.

FEATURES             Location/Qualifiers
     source            1..922
                        /organism="Human orthopneumovirus"
                        /mol_type="genomic RNA"
                        /strain="subgroup A"
                        /isolate="MON 1 87 (Montevideo /Uruguay, 1987)"
                        /db_xref="taxon:11250"
                        /cell_line="HEp-2"
     gene              16..909
                        /gene="G"
     CDS               16..909
                        /gene="G"
                        /codon_start=1
                        /protein_id="CAA83864.1"
                        /db_xref="GOA:Q82078"
                        /db_xref="InterPro:IPR000925"
                        /db_xref="UniProtKB/TrEMBL:Q82078"
                        /translation="MSKTKDQRTAKTLERTWDTLNHLLFISSCLYKLNLSIAQITLS
ILAMIISTSLIIAAIIFIASANKVTLTTAI IQDTSQIKNTTPTYLTQNPQLGISFS
NLSETTSQPTTIPASATPSAESTPQSTTVKTKNTTTTQIQPSKLTTRQKQNKPPNKP
NDFHFEVFNFPVPCISCSNNPTCWAICKRIPNKKPGKTTTKPTKKPTIKTTKKDLKPQ
TTKPKEVPTTKPIEKPTINTTKTNI RTLLTTNTTGNPEHTSQEDTLHSTSSEGNPSP
SQVYTTSEYLSQSPSSSNTTN"
```

ORIGIN

```
1  ggggcaaatg  caaacatgtc  caaaaccaag  gatcaacgca  ccgccaaagc  actagaaaag
61  acctggggaca  ctctcaatca  tctattatcc  atatcatcgt  gcttatatac  gttaaatcct
121  aaatctatag  cacaaatcac  attatccatt  ttggcaatga  taatctcaac  ttcacttata
181  attgcagcca  tcattatcat  agcctcggca  aaccacaaag  tcacactaac  aactgcaatc
241  atacaagata  caacaagcca  gatcaagaac  acaaccccaa  catacctcac  ccagaatccc
301  cagcttgcaa  tcagcttctc  caatctgtcc  gaaactacat  cacaacccac  caccatacca
361  gcctcagcaa  caccaagtgc  tgagtcaacc  ccacaatcca  caacagtcaa  gacaaaaaac
421  acaacaacaa  cccaaataca  acctagcaag  ctcaccacaa  aacaacgcca  aaacaacaca
481  ccaaacacaa  ccaataatga  ttttcacttt  gaagtgttca  actttatacc  ctacagacaa
```

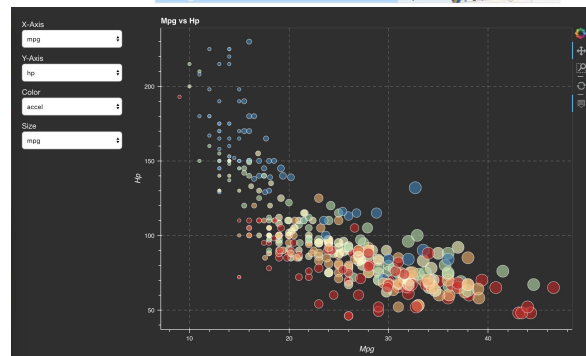
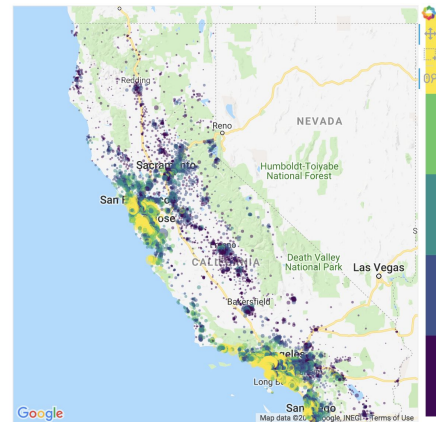
bokeh overview

Pros:

- You can graph a scatter plot on a map
- It's possible to include interactive elements to select how data is visualized
 - i.e. slider to select data range, choose values to represent via dot color or size

Cons:

- Can become slowed down by large datasets
- Does not work well with pandas dataframes



Year released: 1970

End Year released: 2014

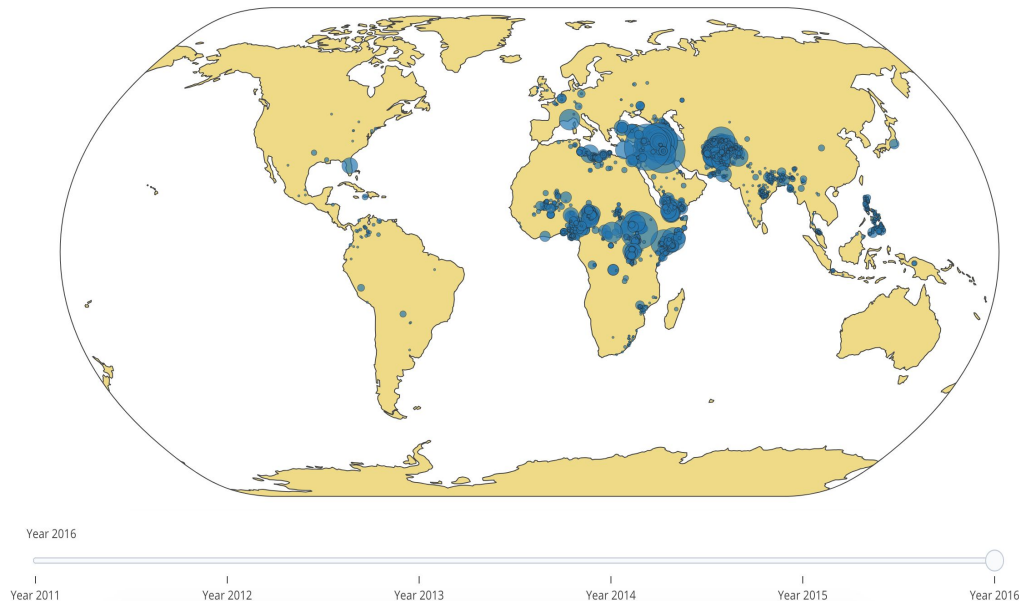
plot.ly overview

Pros:

- Also allows Geo Mapping
- Customizable aesthetics
- Built in interactive features:
 - Toggle to select subsets of data, zoom in/out, easy to download plot
- Many plot types: easy to plot correlations as bar plots without re-formatting data

Cons:

- Time slider is not built-in for maps
- Free version has limit on number of API calls per day



bokeh

vs.

plot.ly

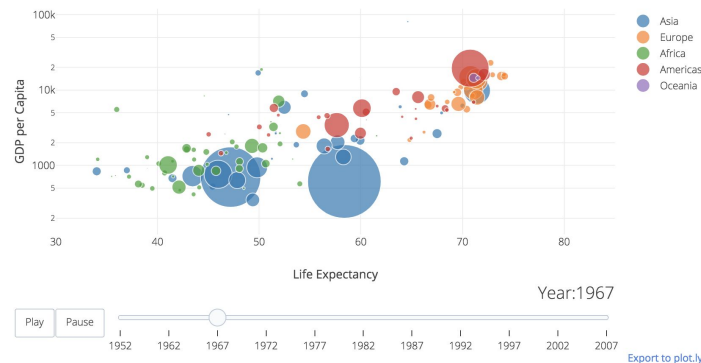
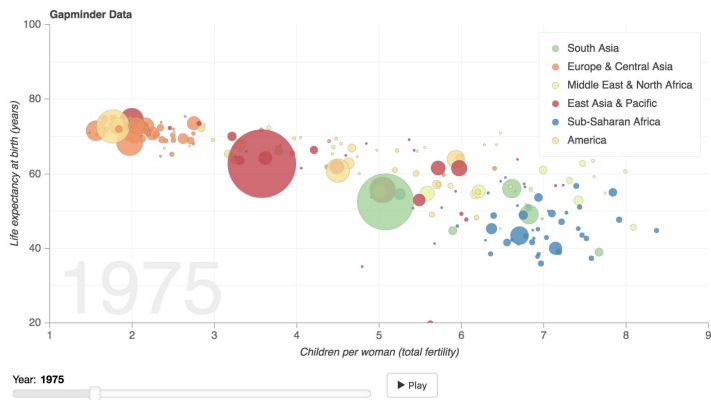
Map RSV sequences based on location collected

Add visual differences for subtype and number of sequences

- Similar functionalities
 - Can overlay on Google Maps
- Similar functionalities
 - Free version may have color limitations

Interactive map

- Often options look more polished
- Easy to add sliders and animate



bokeh

vs.

plot.ly

Overlay additional data on map (use colors, toggles, insets, etc.)

- We will need to try several visualization options to determine what looks best, but both packages have several options.

OVERALL

- | | |
|--|--|
| <ul style="list-style-type: none">• More complicated to implement• Pandas DataFrames must be converted to bokeh objects• Lots of documentation, but not necessarily for our use case | <ul style="list-style-type: none">• Easy to implement, all in python• Works well with pandas• Lots of documentation with examples very similar to our use case |
|--|--|

Due largely to ease of use, we are planning on using **plot.ly** for our visualizations.

