Qualitative Discovery of Marine Population Trends

Student: Andrea Baisero
Community Partner: Benjamin Moran

Overview

Problem Description

Design

Conclusions

Problem Description

Problem Description

Project Partner:

Benjamin Moran, undergraduate Senior in Marine Biology.

Data Description:

Sample statistics on various fish species obtained over the course of 20 years in 3 regions: the Florida Keys, Dry Tortugas National Park, Southeast Florida Coral Reef Initiative (SEFCRI) areas.

Datasets

Taxonomy: Biological hierarchy of family, genera, and species.

Samples: Histogram:

- Region, i.e. (FLA KEYS, DRY TORT, SEFCRI);
- Date;
- Location, i.e. latitude and longitude;
- Size, (sometimes negative!!!);
- Depth;
- Protection Status, (location dependent);
- Count, (not always integer!!!);

Strata: Info on sampling areas, e.g. grid sizes, prot. status

Notes

1) Very big (samples.csv $\sim$1GB)

2) Infested with missing fields, *semantically wrong values* (!!!).

3) Original data server has been offline for various weeks.

Problem Description

Project Partner:

Benjamin Moran, undergraduate Senior in Marine Biology.

Visualization Needs:

Discover explanatory conditions for variations in sample statistics.

Dependent variables: size and abundance.

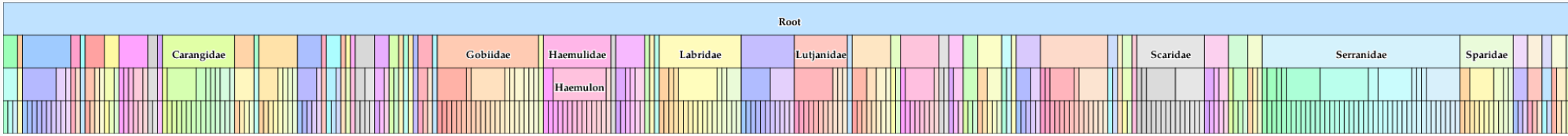
Independent variables: location, date, protection status, depth, distance from shore, and habitat type.

Design

Design Goals

- Visualize sample data on top of a geographic map.
- Filter data according to region, taxonomy, location, date, protection status, ...
  $\Rightarrow$ Visualize, navigate, and select elements of taxonomy tree.
- In-detail views specific to selected filters, with appropriate within-view and inter-view comparison capabilities.
- Preliminary analysis via standard ML techniques (e.g. KDE for data density estimation, GP for data regression lines)

Design Overview



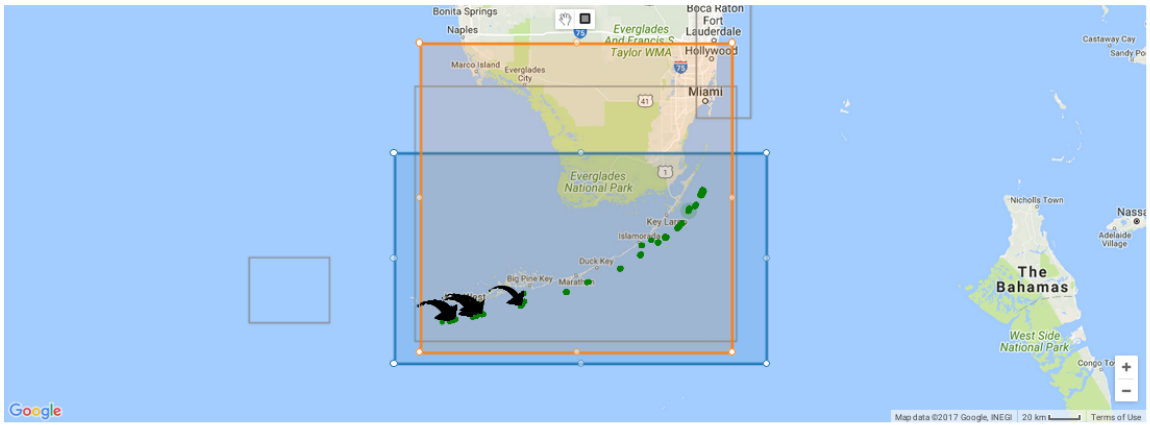
Sample View Controls

X Axis: date
Y Axis: length [in]
Switch Axes: X <---> Y

KDE Kernel: Gaussian
KDE Scale:

GP Kernel: Periodic
GP Scale:
GP Period: 1 Year

Axes-Origin:
Linked-Views:
Views-Average:



Geo-Time View Controls

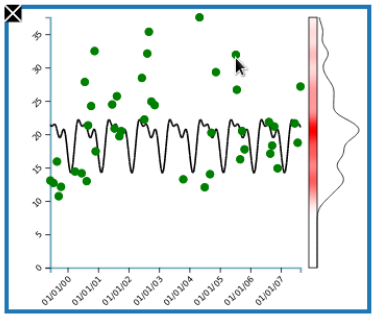
Region: FLA KEYS
P. Status: Protected

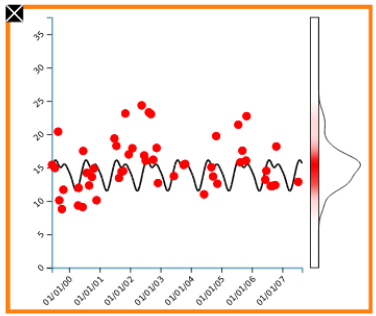
N. Samples: 3172
N. Creatures: 2219

Data Weighted Average

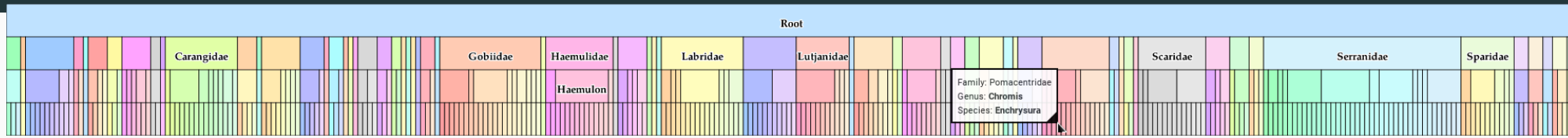
Family: Serranidae
Date: 7/11/2005
Length: 31.92
Number: 53
Depth: 10.79
Lat: 24.47
Lng: -81.77

Dec 1998
Dec 2017





Taxonomy Tree

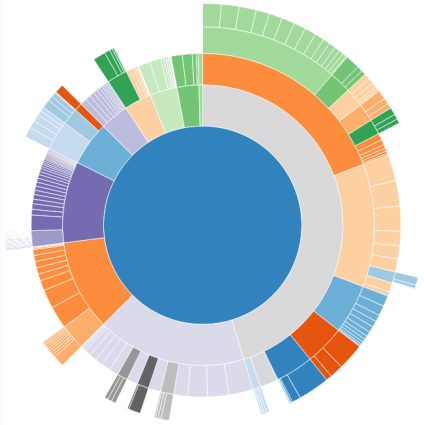


Goal: Show, navigate and select Family, Genus or Species.

Difficulties: Shallow, but high branching factor.

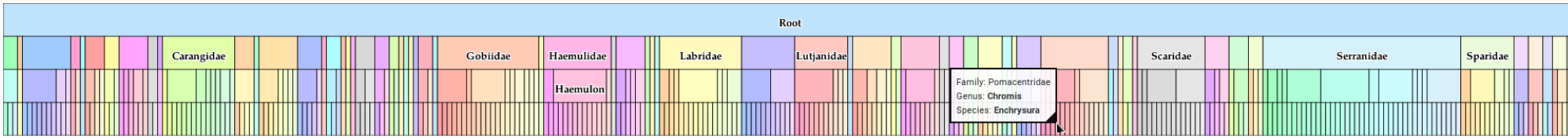
Key concern: space efficiency.

Solution: Inspired by Zoomable Sunburst[1], with key domain-specific changes.



[1] *Zoomable Sunburst*. `https://bl.ocks.org/mbostock/4348373`.

Taxonomy Tree – Visualization

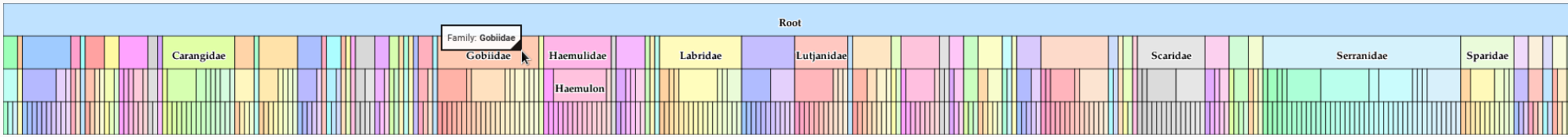


Domain-specific features:

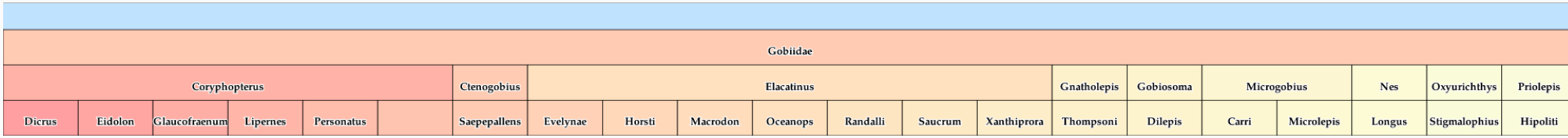
- Flattened out to exploit horizontal screen extension.
- Ordered hierarchically/alphabetically, to aid search.
- Labels only shown if sufficient space is available.
  (More details available via tooltip + navigation)
- Palette chosen to aide distinction between families
  (small variations in hue applied to descendant genera/species)

Taxonomy Tree – Navigation and Selection

Clicking on an element. . .



. . . expands it, revealing new labels and allowing for tree navigation.



Ctrl + click selects a family/genus/species for filtering purposes.

Geographic View

Goal: Show, navigate and select by regions and area

Key Concern: Simple navigation (zooming, panning).

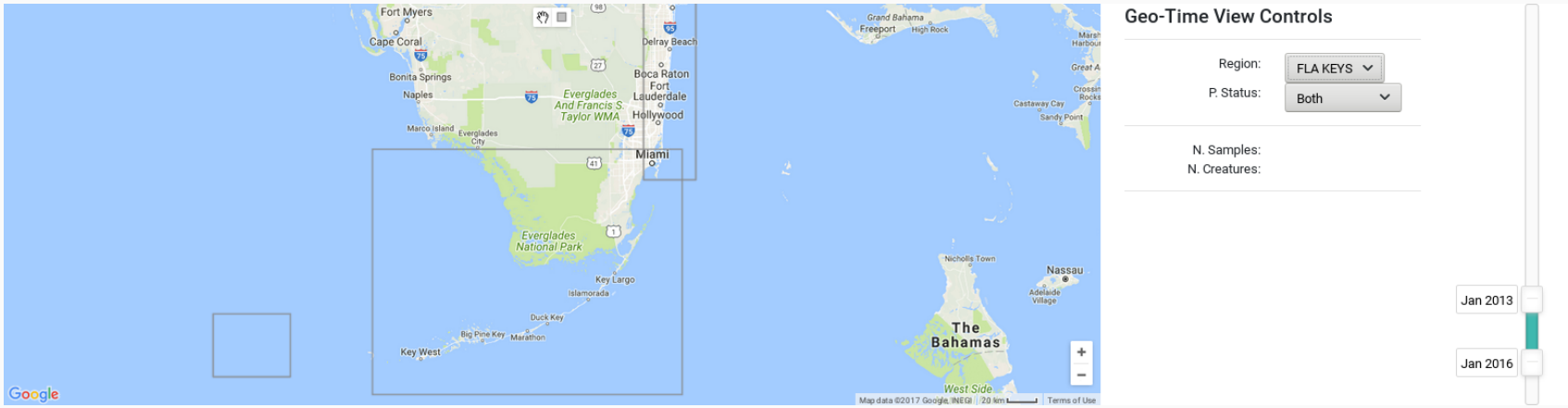
Solution: Based on Google Maps[2].

Pros: Many built-in functionalities (zooming, panning, coordinates-to-pixel mappings).

Cons: Somewhat slow, and JS-blocking!!!

[2] *Google Maps JavaScript API*. `https://developers.google.com/maps/documentation/javascript/`.

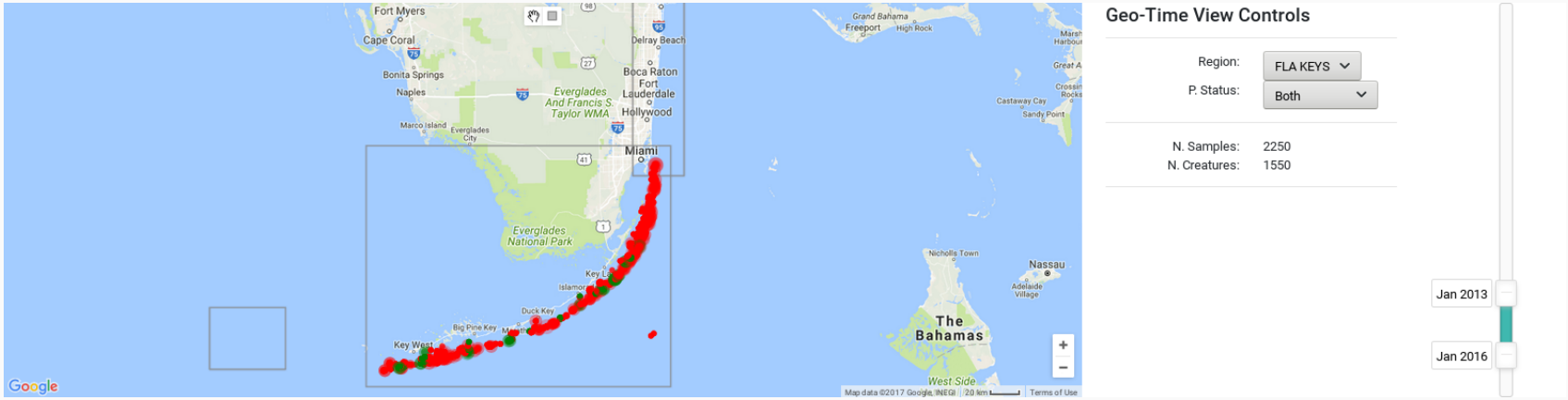
Geographic View – Elements



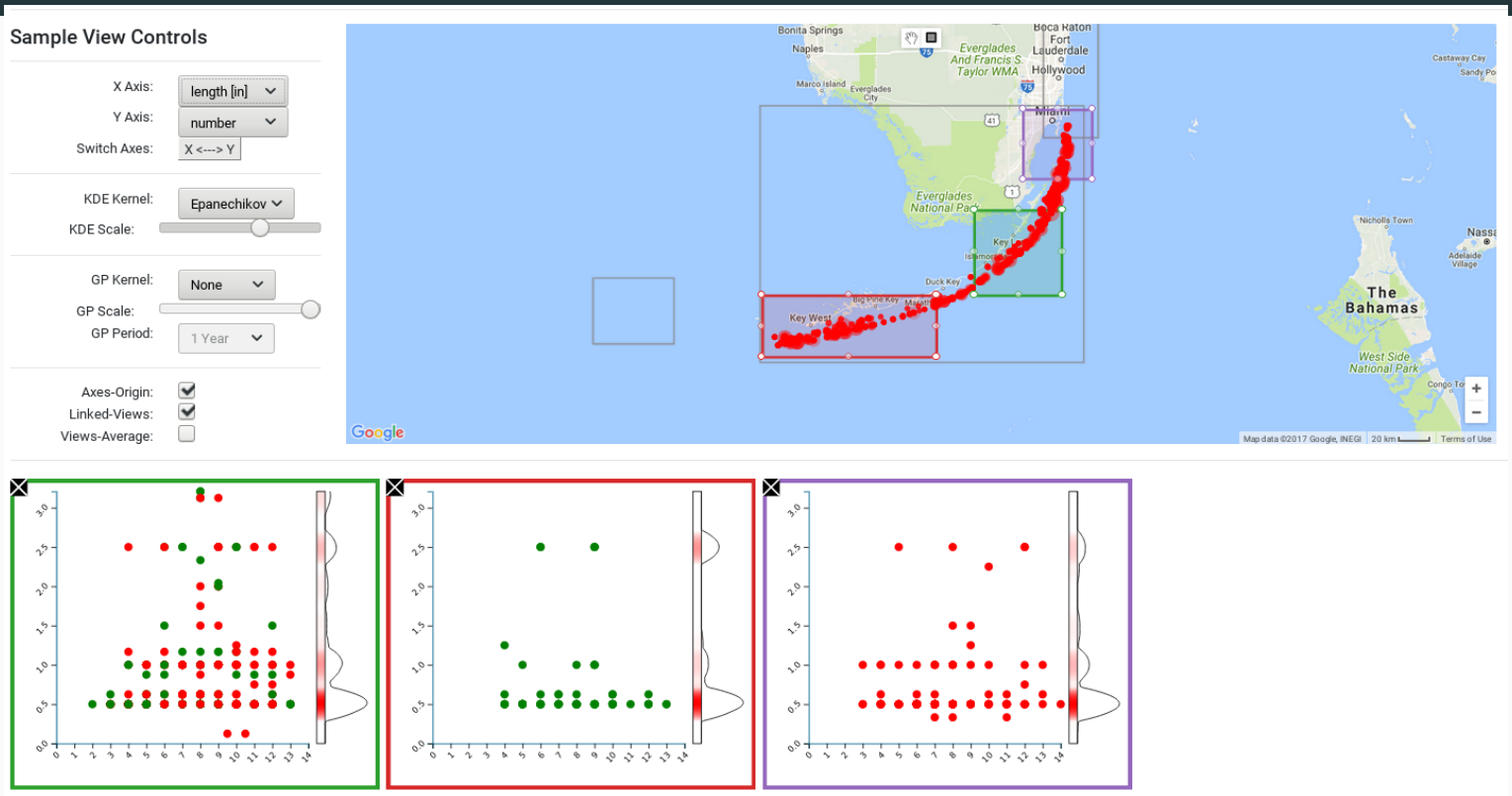
Key Elements

- Map
- Date Slider
- Region drop-down selection
- P. Status drop-down selection
- Empty space for labels + hover info

Geographic View – Examples



Detail Views

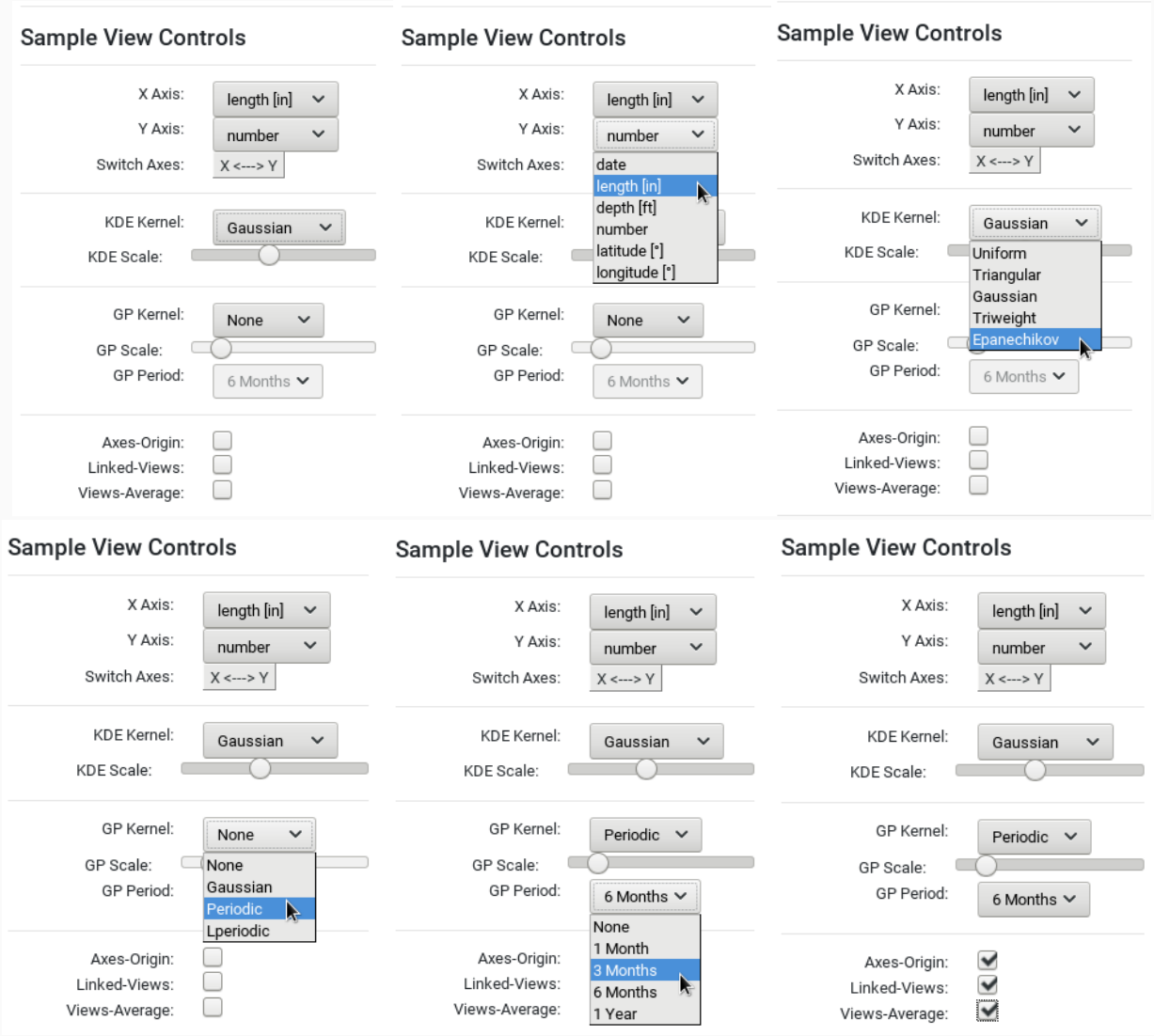


Goal: Maximum flexibility in view manipulation, (similar to "From Detail to Overview" and "Glo-stix").

Key Concerns: Maintaining data persistency.

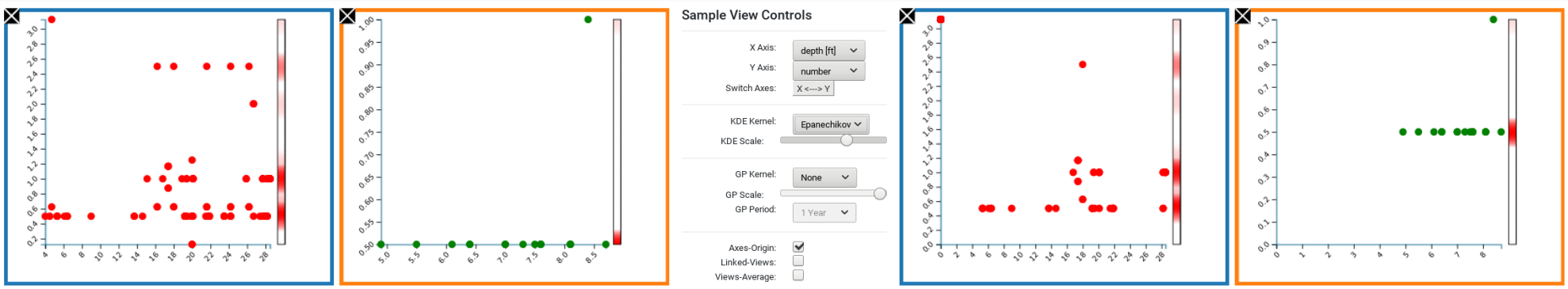
Ability to explore all sorts of potential relationships.

Detail Views – Controls

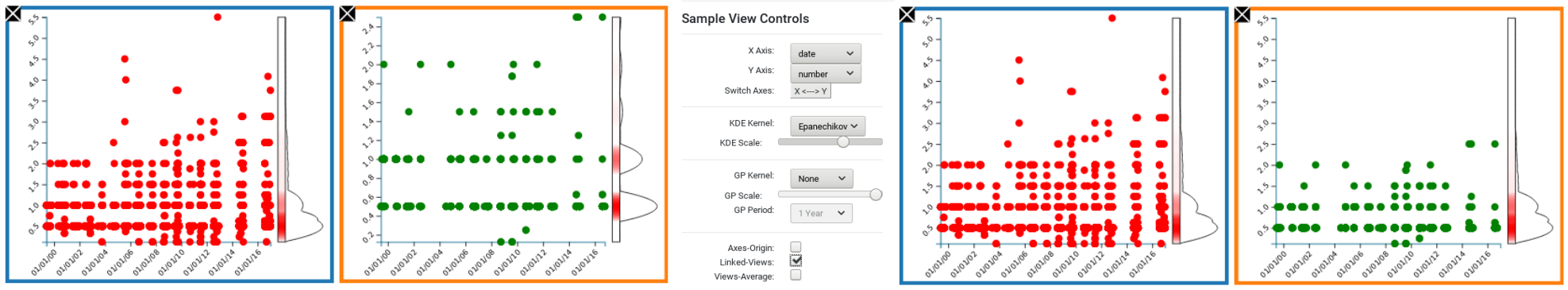


Detail Views – Avoiding Standard Plotting Biases

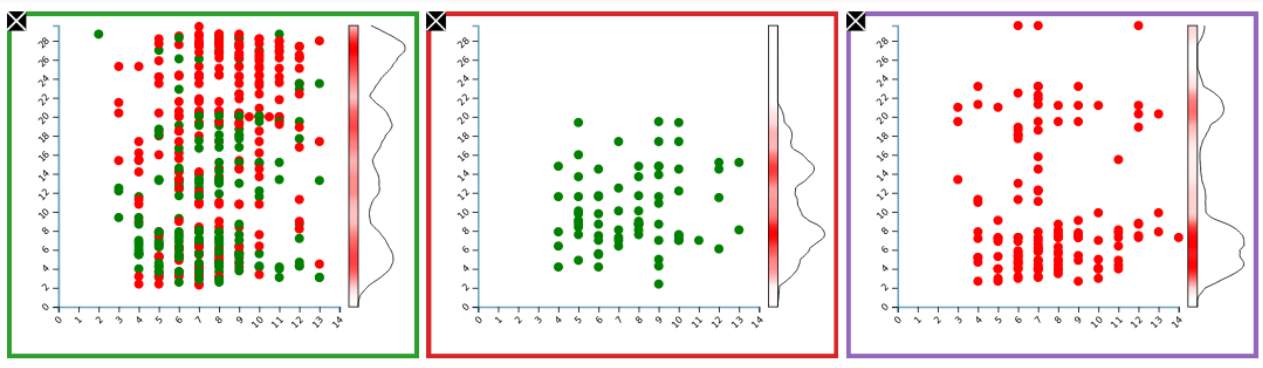
Axis Origin:

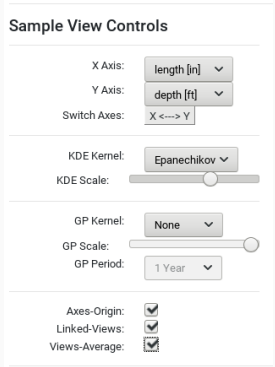


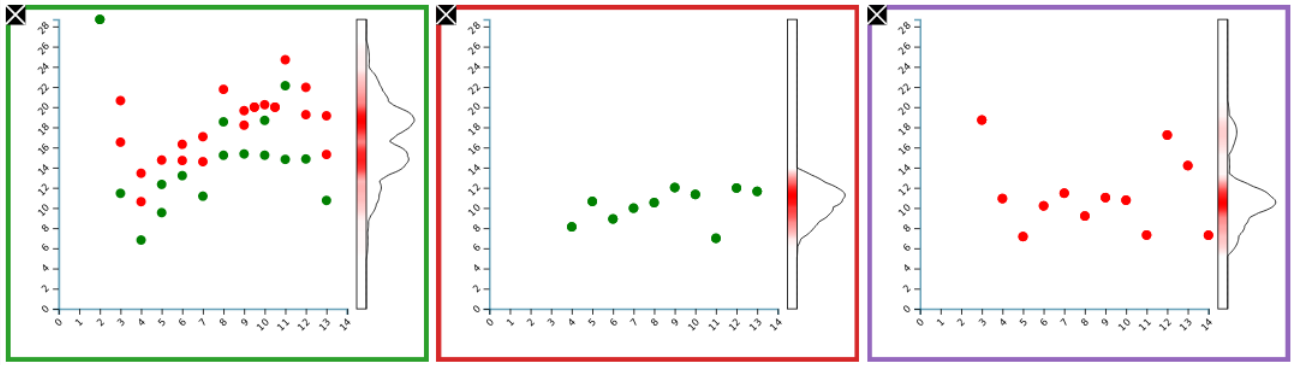
Linked Axes:



Detail Views – Averaging over Sample Y-Axis values

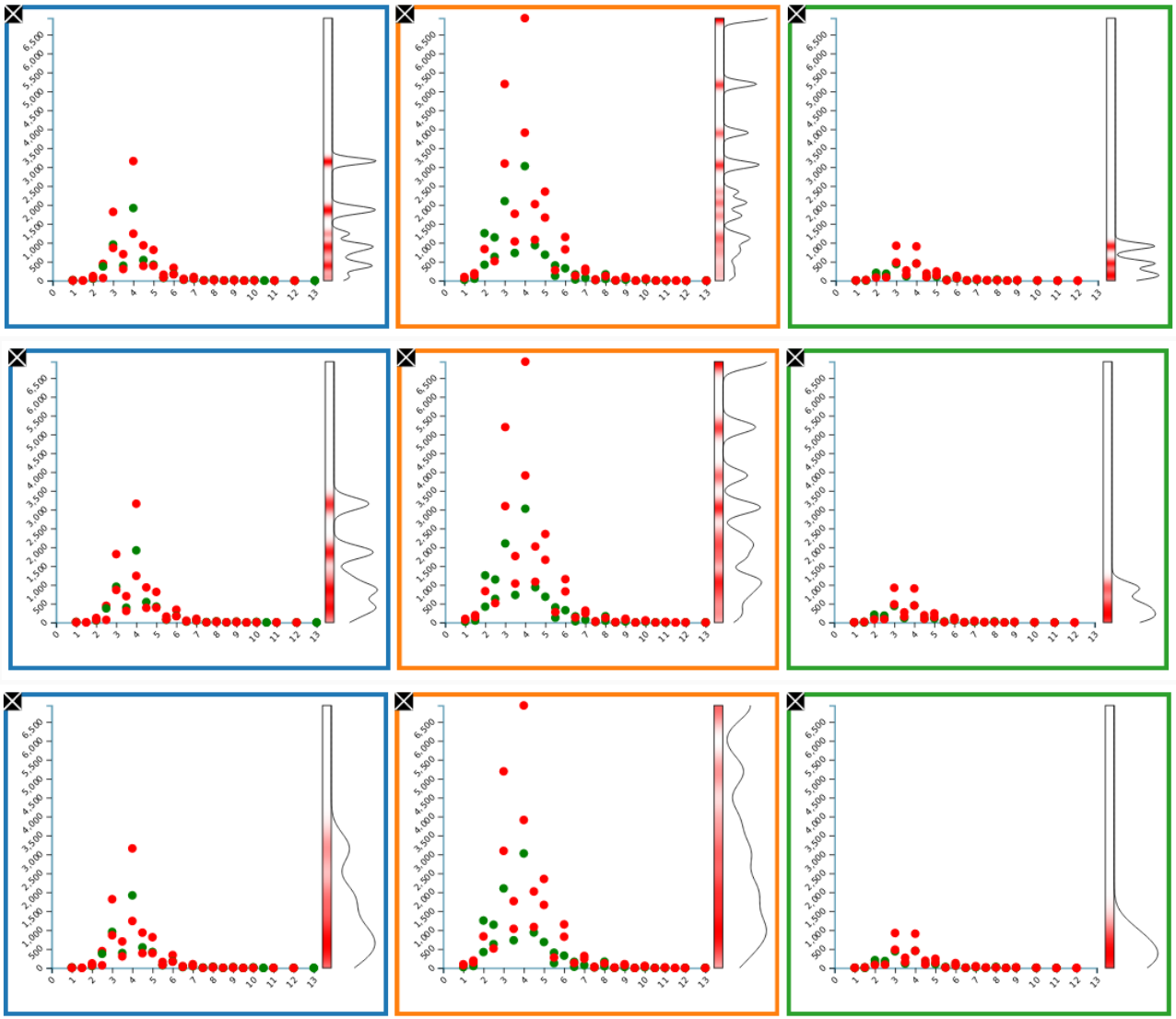






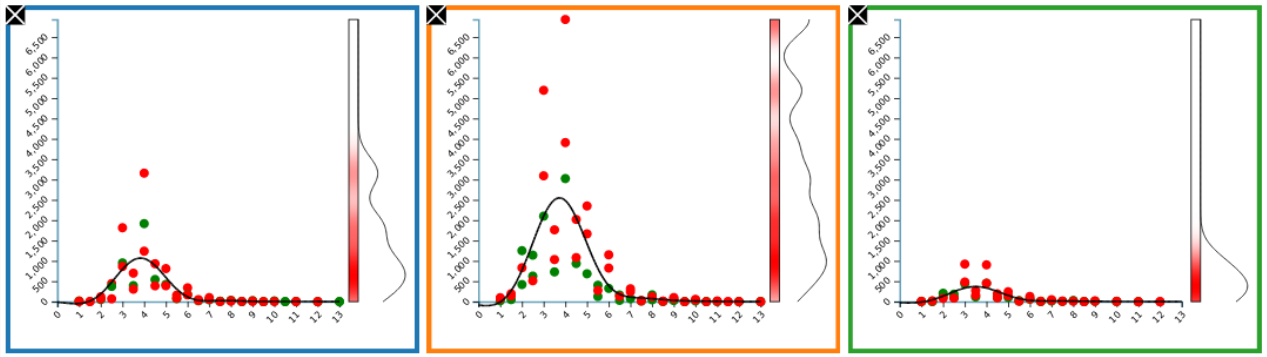
Detail Views – KDE

Gaussian Kernel (varying length-scales):

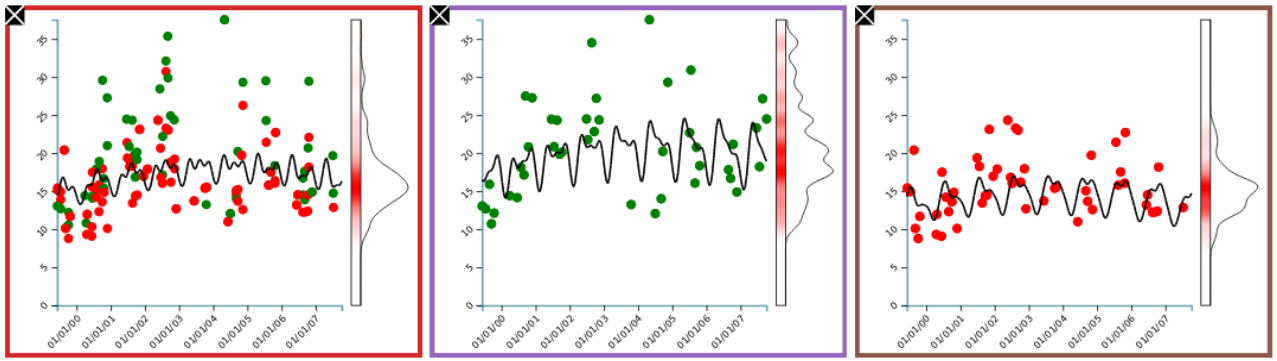


Detail Views – GP Regression

Gaussian Kernel:



Locally Periodic Kernel:



Live Demo

Conclusions

Conclusions

Achieved Design Goals

- Efficient taxonomy navigation and selection.
- Axis manipulation aids exploration of relationships within- and between- detail view.
- Basic ML analysis.

Future Work

- Improve data fields and data server efficiency.
- Switch from Google Maps to lighter d3-geo solution.
- Extend data-dim / axis manipulation.
- Better management of combined filters.

Questions ?

References

- *Google Maps JavaScript API*. `https://developers.google.com/maps/documentation/javascript/`.
- Stolper, Charles D et al. "Glo-stix: Graph-level operations for specifying techniques and interactive exploration". In: *IEEE transactions on visualization and computer graphics* 20.12 (2014), pp. 2320–2328.
- Van den Elzen, Stef and Jarke J Van Wijk. "Multivariate network exploration and presentation: From detail to overview via selections and aggregations". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2310–2319.
- *Zoomable Sunburst*. `https://bl.ocks.org/mbostock/4348373`.