

# Visualization 1

## What can we learn from the visualization?

To get an overview of the data by understanding the distribution of the recorded coordinates, any identified points of interest and their relative location on the world map as we know it today.

## What is the name for the type of visualization(s) used?

Map

```
import altair as alt
from vega_datasets import data

places = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/pleiades_places.json')
graticule = alt.topo_feature(data.world_110m.url, 'countries')

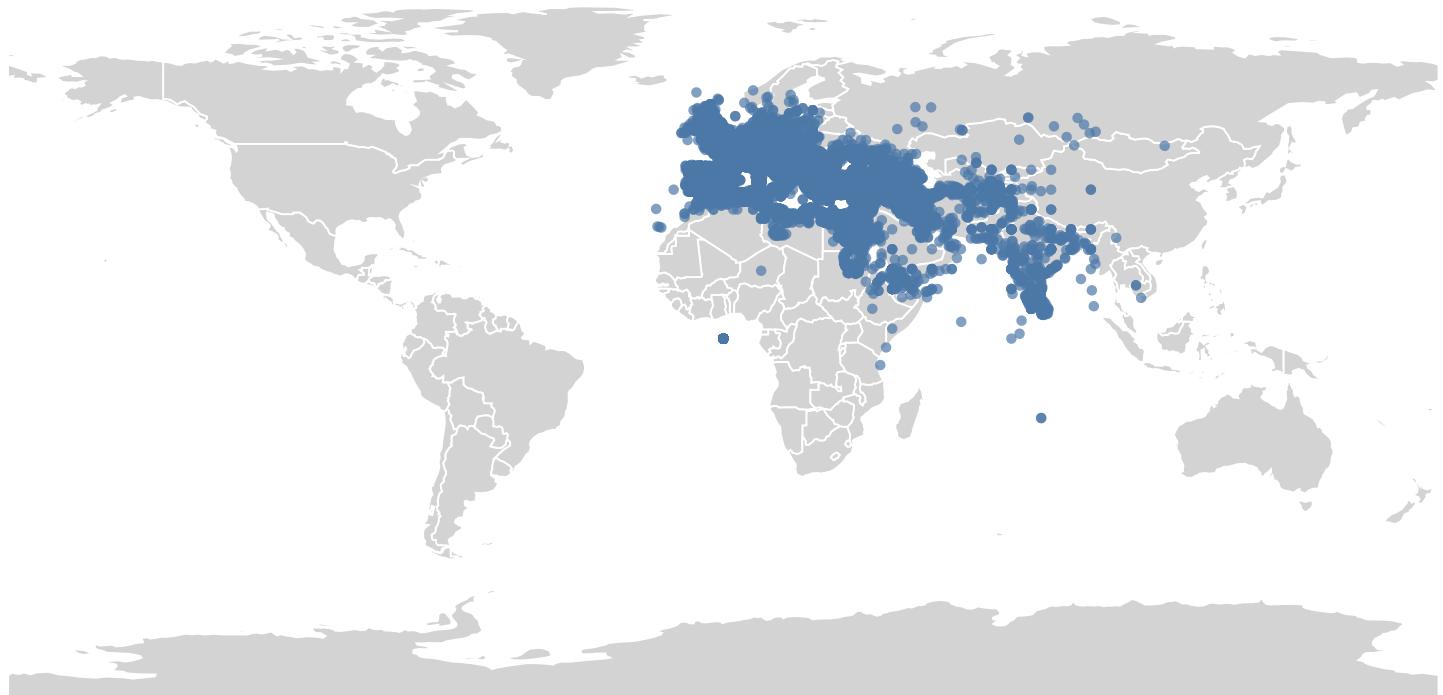
featureTypesSelector = alt.selection_single(empty='all', fields=['featureTypes'])

mapBase = alt.Chart(graticule).mark_geoshape(
    fill='lightgray',
    stroke='white'
).project(
    'equirectangular'
).properties(
    width=750,
    height=500
)

placePoints = alt.Chart(places).mark_circle().encode(
    latitude='reprLat:Q',
    longitude='reprLong:Q',
    tooltip=['title:N', 'featureTypes:N']
).transform_filter(
    featureTypesSelector
).properties(
    title="Identified coordinates for the pleiades places dataset"
).add_selection(featureTypesSelector)

(mapBase + placePoints).configure_view(stroke=None)
```

## Identified coordinates for the pleiades places dataset



### What are all visual mappings used?

#### *latitude*

reprLat

#### *longitude*

reprLong

#### *tooltip*

title, featureTypes

### Was there any special data preparation done?

The data is dynamically filtered such that when a data point on the map is clicked, only data points that are of the same featureType as the point clicked are rendered on the map. Clicking on any space on the map besides a data point resets the filter.

### What are the limitations of your design?

More interaction with the map such as the ability to pan and zoom would be beneficial. Further, being able to show the connections between places might help provide more context and, thus, a better understanding of the dataset.

# Visualization 2

**What can we learn from the visualization?**

To understand the correlation between the latitude and longitude coordinates across each pleiades dataset and the precision of the coordinates

**What is the name for the type of visualization(s) used?**

Dot Dash Plot with linked histogram

```

import altair as alt

locations = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/locations.csv')
names = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/pleiaades.csv')
places = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/plexippus.csv')

brush = alt.selection_interval()
tick_axis = alt.Axis(labels=False, domain=False, ticks=False)

def scatter(source):
    return (
        alt.Chart(source).mark_circle(opacity=0.75).encode(
            x='reprLat:Q',
            y='reprLong:Q',
            color=alt.condition(brush, 'locationPrecision:N', alt.value('lightgray')),
            tooltip=['title:N', 'reprLatLong:N', 'minDate:N', 'maxDate:N']
        ).add_selection(brush)
    )

def xTicks(source):
    return alt.Chart(source).mark_tick().encode(
        alt.X('reprLat:Q', axis=tick_axis),
        alt.Y('locationPrecision:N', title='', axis=tick_axis),
        color=alt.condition(brush, 'locationPrecision:N', alt.value('lightgrey')),
    ).add_selection(brush)

def yTicks(source):
    return alt.Chart(source).mark_tick().encode(
        alt.X('locationPrecision:N', title='', axis=tick_axis),
        alt.Y('reprLong:Q', axis=tick_axis),
        color=alt.condition(brush, 'locationPrecision:N', alt.value('lightgrey'))
    ).add_selection(brush)

def histo(source):
    return (
        alt.Chart(source).mark_bar().encode(
            x='locationPrecision:N',
            y='count(locationPrecision):Q',
            color='locationPrecision:N',
            tooltip=['count(locationPrecision):N']
        ).transform_filter(
            brush
        ).interactive()
    )

def dotDashPlot(sources):
    return yTicks(sources) | (scatter(sources) & xTicks(sources)) | histo(sources)

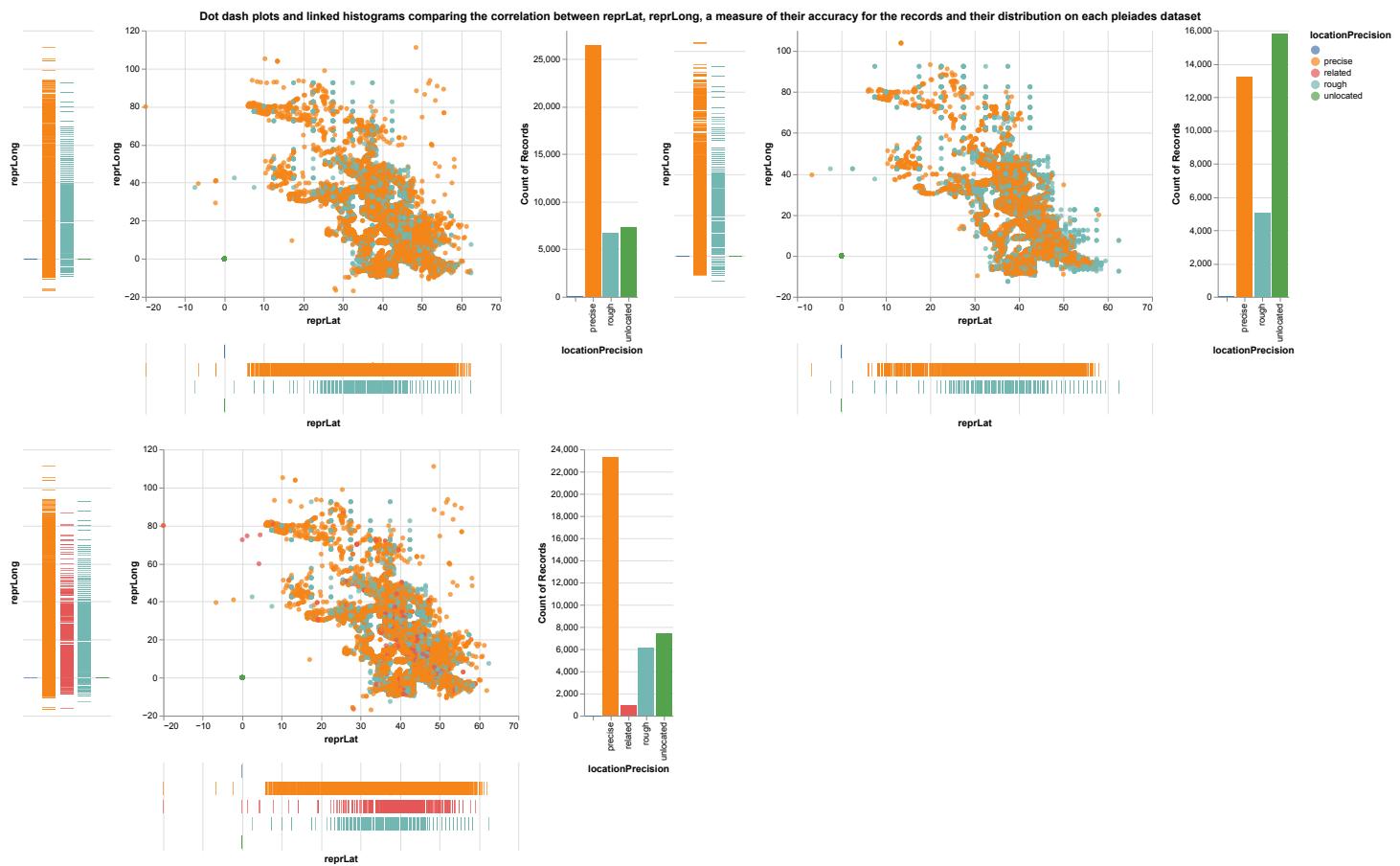
alt.vconcat(
    (dotDashPlot(locations) | dotDashPlot(names)),
    dotDashPlot(places)
)

```

```

).properties(
    title='Dot dash plots and linked histograms comparing the correlation between reprLat, reprLong, a measure of their accuracy for the records and their distribution on each pleiades dataset'
).configure_title(orient='top', anchor='middle').configure_view(stroke=None)

```



## What are all visual mappings used?

For each pleiades dataset,

### scatter plot

**x position:** `reprLat`

**y position:** `reprLong`

**color:** conditionally `locationPrecision` or lightgray

**tooltip:** title, `reprLatLong`, `minDate`, `maxDate`

### x-axis strip plot

**x position:** `reprLat`

**y position:** locationPrecision

**color:** conditionally locationPrecision or lightgray

#### *y-axis strip plot*

**x position:** locationPrecision

**y position:** reprLong

**color:** conditionally locationPrecision or lightgray

#### *histogram*

**x position:** locationPrecision

**y position:** count of locationPrecision records

**color:** locationPrecision

**tooltip:** count of locationPrecision records

### **Was there any special data preparation done?**

Linked brushing is set up such that when an area on a scatter plot is highlighted, corresponding areas on the other scatter plots are highlighted and the connected strip plot and histogram for each are dynamically updated. Clicking on any space on any scatter plot besides a data point resets the visualization.

### **What are the limitations of your design?**

The visualization could be improved by adding input elements such as checkboxes, dropdowns and/or radio buttons to allow the ability to choose what coordinates to display based on a selected precision. A line of best fit might also be beneficial in understanding the correlation between the latitude and longitude data channels.

# Visualization 3

**What can we learn from the visualization?**

To understand the correlation between the minDate and maxDate for the records across each pleiades dataset

**What is the name for the type of visualization(s) used?**

Binned bubble plot and histogram

```

import altair as alt

locations = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/locations')
names = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/names')
places = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/places')

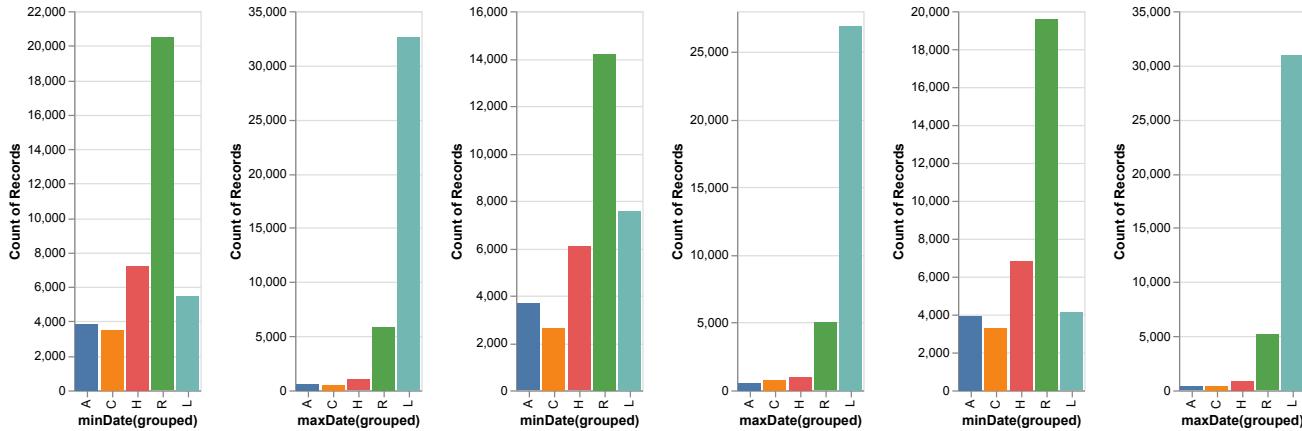
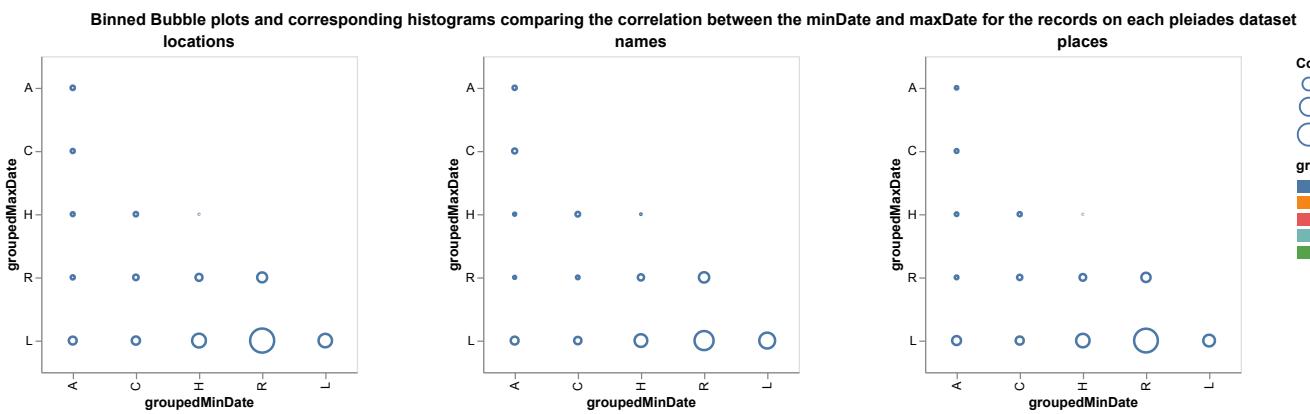
def binnedBubblePlot(title, source):
    return (
        alt.Chart(source).transform_calculate(
            groupedMinDate='datum.minDate < -550 ? "A" : datum.minDate < -330 ? "C" : datum.minDate >= -330 ? "R" : "L"',
            groupedMaxDate='datum.maxDate < -550 ? "A" : datum.maxDate < -330 ? "C" : datum.maxDate >= -330 ? "R" : "L"'
        ).transform_calculate(
            groupedCount='count()'
        ).mark_point().encode(
            alt.X("groupedMinDate:0", sort=['A', 'C', 'H', 'R', 'L']),
            alt.Y("groupedMaxDate:0", sort=['A', 'C', 'H', 'R', 'L']),
            size='groupedCount:N',
            tooltip=['groupedCount']
        ).properties(width=250, height=250, title=f'{title}')
    )

def histo(source, column):
    return (
        alt.Chart(source).transform_calculate(
            groupedDate=f'datum.{column} < -550 ? "A" : datum.{column} < -330 ? "C" : datum.{column} >= -330 ? "R" : "L"'
        ).mark_bar().encode(
            alt.X("groupedDate:0", sort=['A', 'C', 'H', 'R', 'L'], title=f'{column}(groupedDate)'),
            alt.Y('count()'),
            color='groupedDate:N',
            tooltip=['groupedDate:N', 'count()']
        ).interactive()
    )

def combi(title, source):
    return (
        binnedBubblePlot(title, source) & (histo(source, 'minDate') | histo(source, 'maxDate'))
    )

alt.hconcat(
    combi('locations', locations), combi('names', names), combi('places', places)
).properties(
    title='Binned Bubble plots and corresponding histograms comparing the correlation between locations, names and places'
).configure_title(orient='top', anchor='middle')

```



## What are all visual mappings used?

For each pleiades dataset,

### *binned bubble plot*

**x position:** minDate (grouped and sorted according to the group)

**y position:** maxDate (grouped and sorted according to the group)

**tooltip:** count of records

Given that date is minDate or maxDate

### *histogram*

**x position:** date (grouped and sorted according to the group)

**y position:** count of records

**color:** date (grouped and sorted according to the group)

**tooltip:** count of records

## Was there any special data preparation done?

For each dataset, the data has been grouped to match the pleiades README on `timePeriods` such that the records fall into the following bins "... 'A' (1000-550 BC), 'C' (550-330 BC), 'H' (330-30 BC), 'R' (AD 30-300), 'L' (AD 300-640)".

## What are the limitations of your design?

The size scale could be more carefully chosen for better visibility of all bubbles, especially those on the smaller end of the spectrum. Then, for easier, quicker comparisons between each chart, each bubble could be overlay with the count of records it represent as opposed to relying for tooltips. The charts could also be organised and sized such that for each bubble plot, the corresponding `minDate` histogram is below it and the `maxDate` histogram is 'turned' on its left side and places to the left of the bubble plot and the scales line up.

# Visualization 4

## What can we learn from the visualization?

To understand the distribution of the `timePeriods` within which points of interest were founded and dissolved.

## What is the name for the type of visualization(s) used?

Heatmap

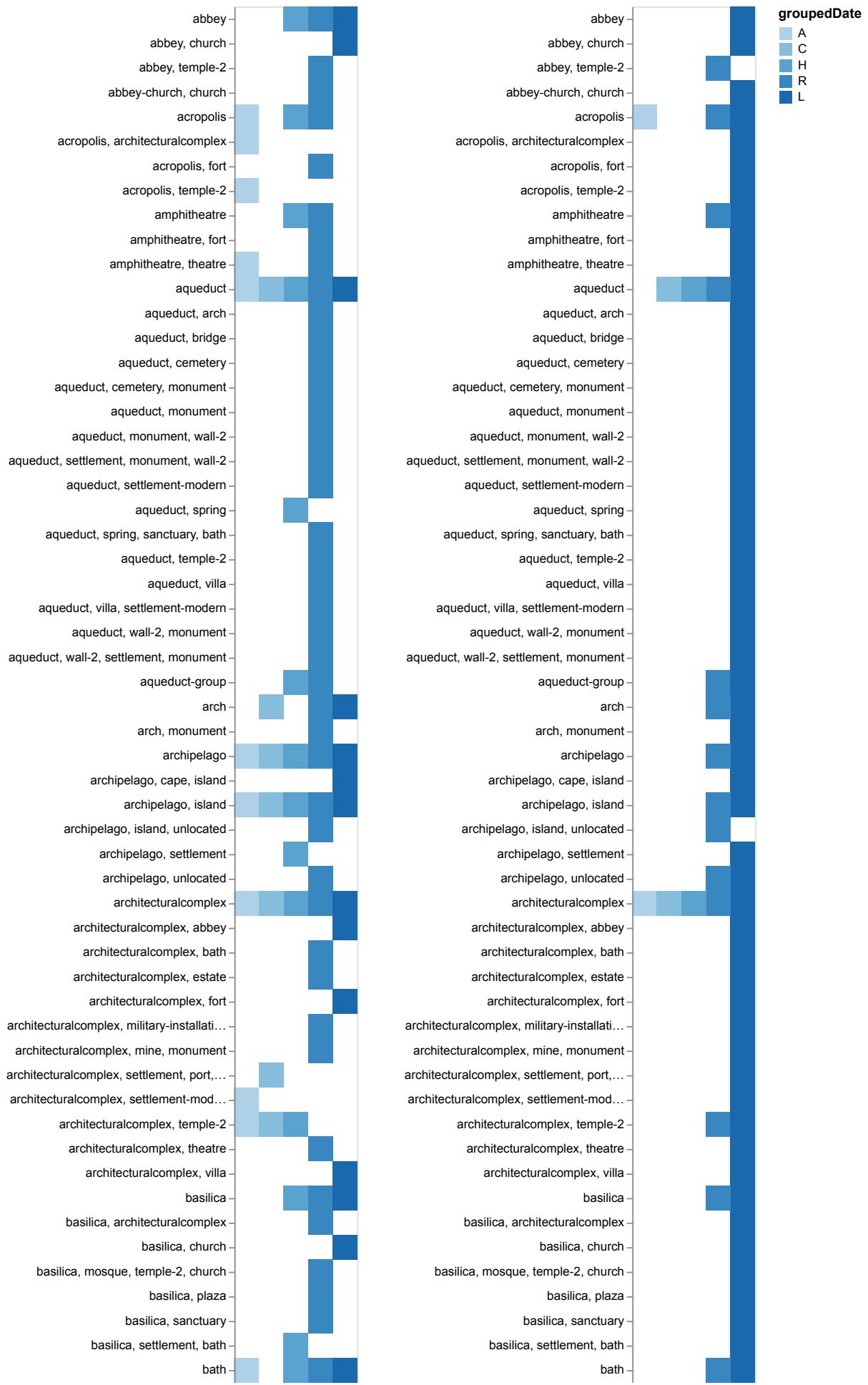
```
import altair as alt

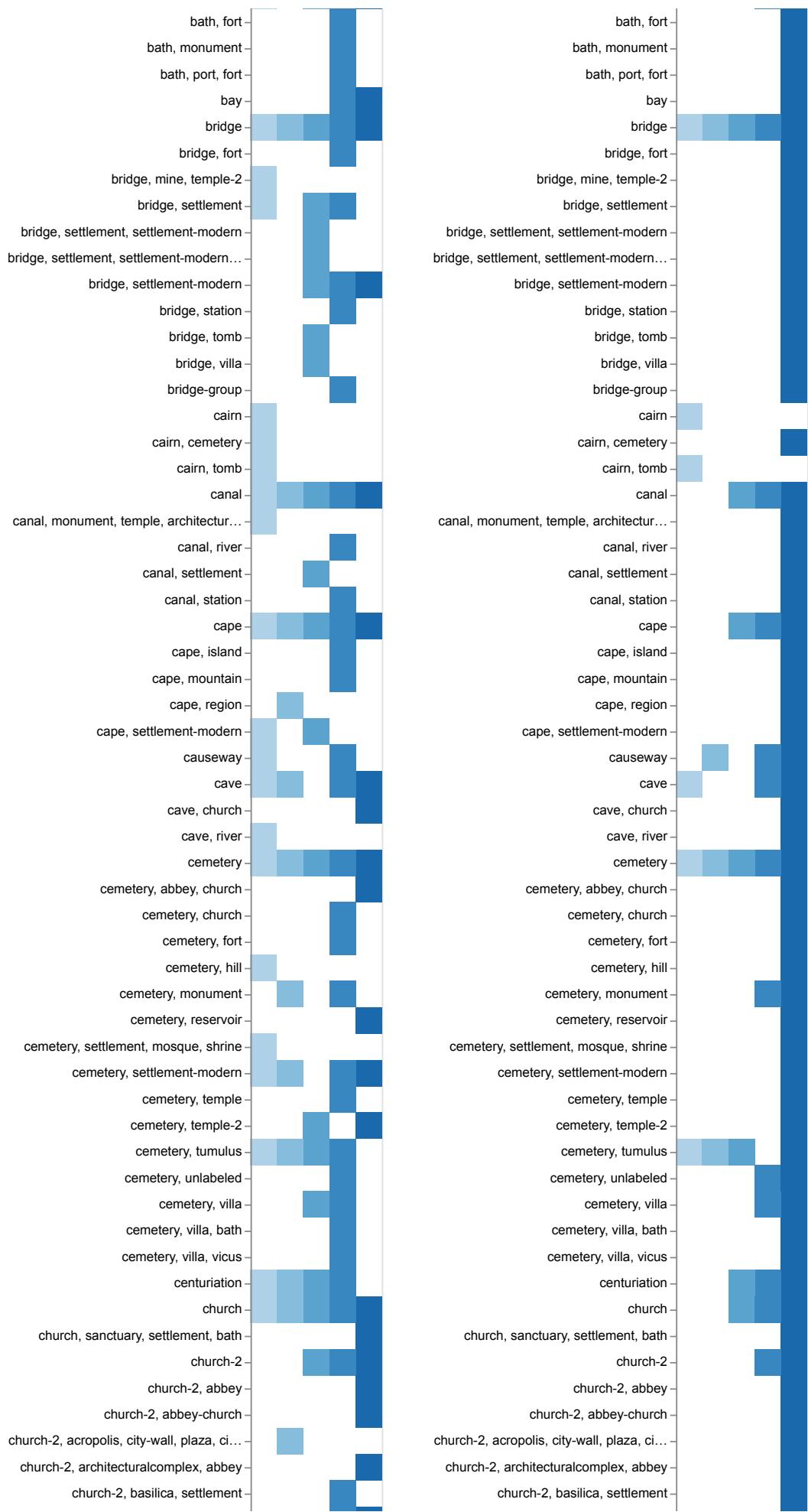
places = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/ple

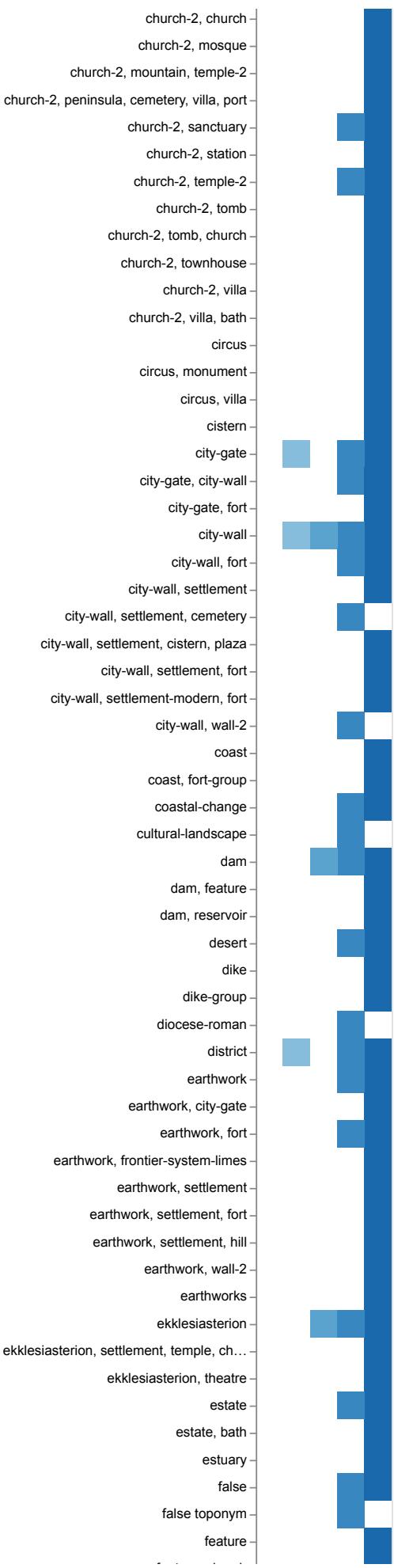
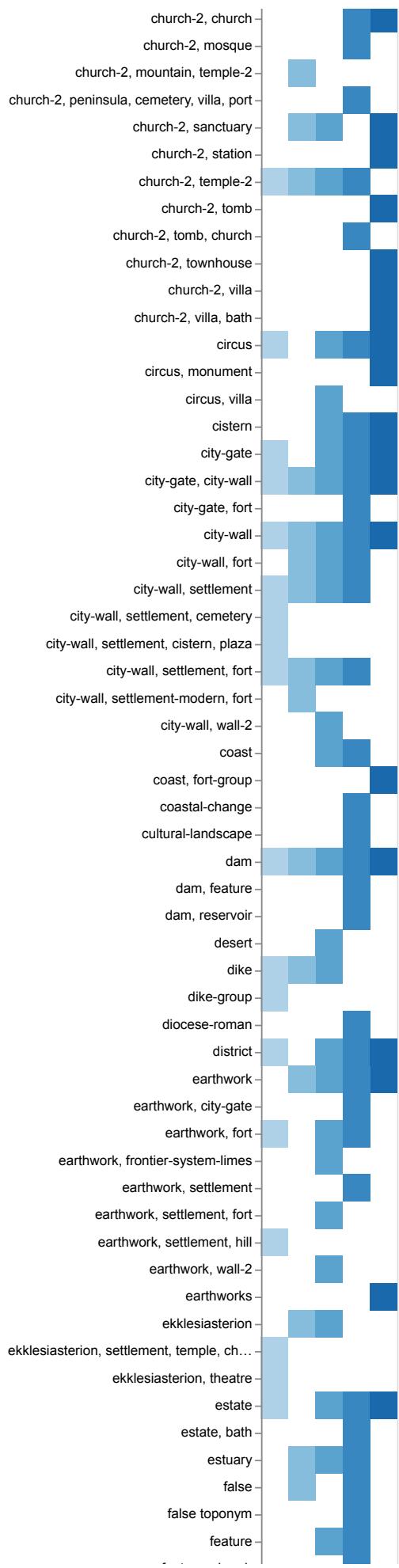
def heatMap(column):
    return (
        alt.Chart(places).mark_rect().transform_calculate(
            groupedDate=f'datum.{column} < -550 ? "A" : datum.{column} < -330 ? "C" : d
        ).encode(
            alt.X("groupedDate:0", sort=['A', 'C', 'H', 'R', 'L'], title=f'{column}(gro
            y='featureTypes:N',
            color=alt.Color('groupedDate:0', sort=['A', 'C', 'H', 'R', 'L']),
            tooltip=['count()']
        )
    )

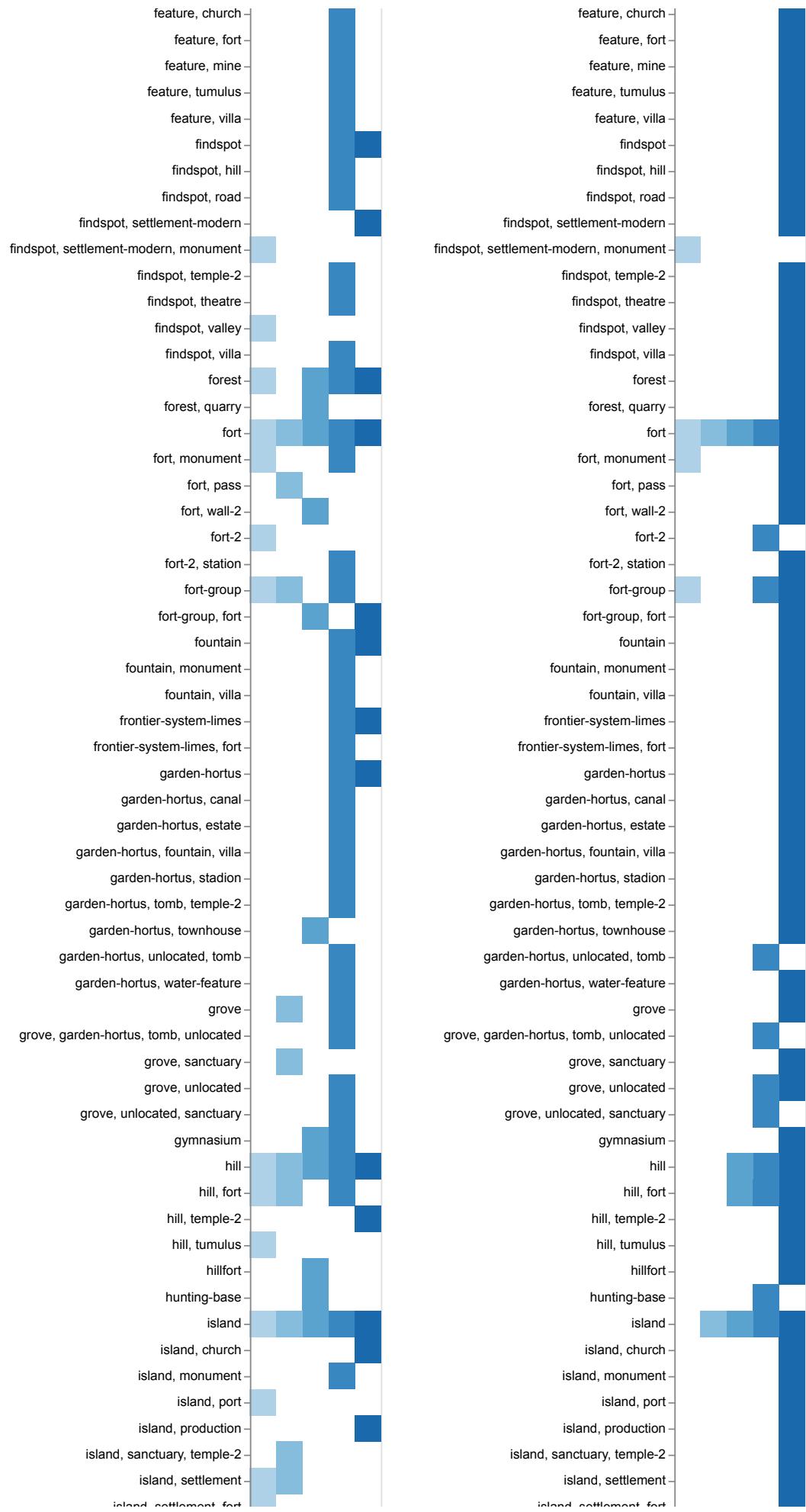
alt.hconcat(
    (heatMap('minDate') | heatMap('maxDate'))
).properties(
    title='Heatmaps showing what featureTypes have a minDate and maxDate in what timePe
).configure_title(orient='top', anchor='middle')
```

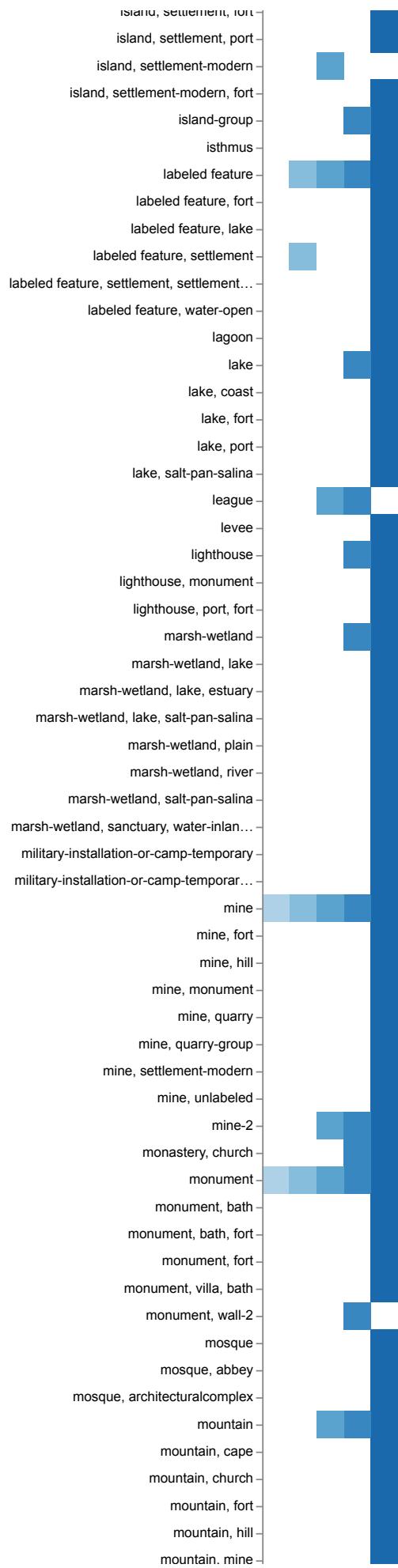
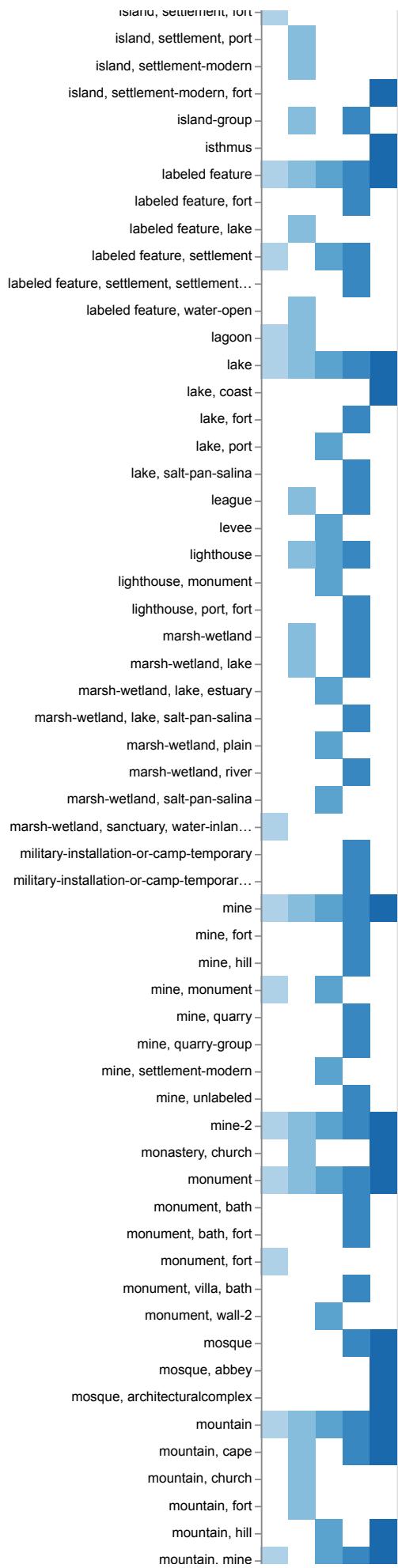
Heatmaps showing what featureTypes have a minDate and maxDate in what timePeriods for each record on the places pleiades dataset

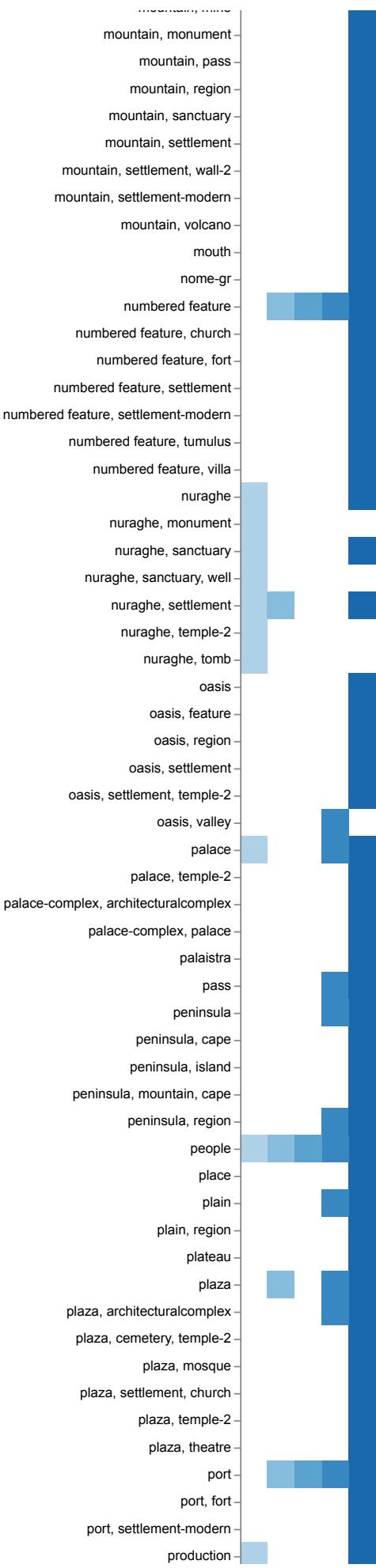
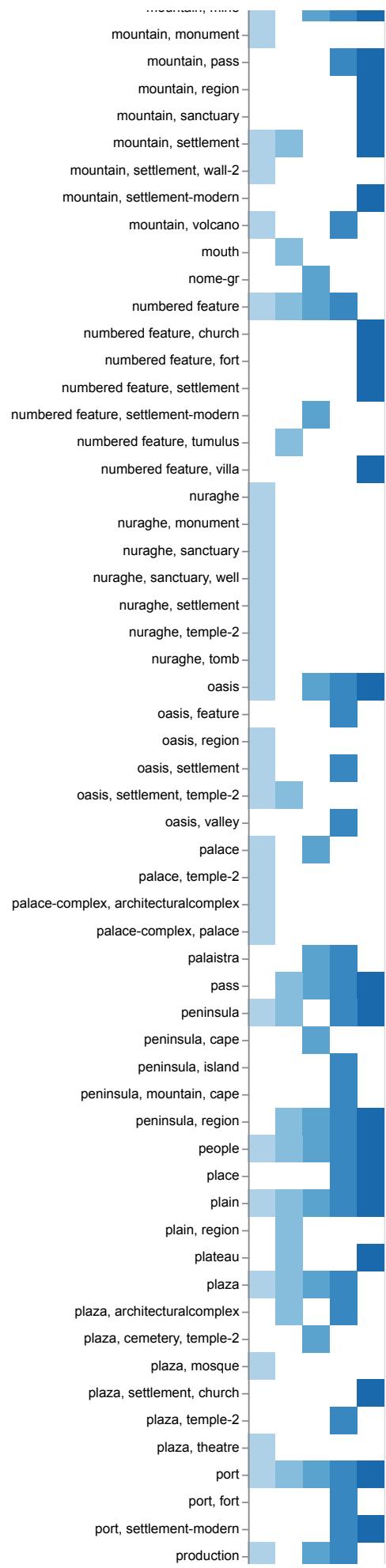


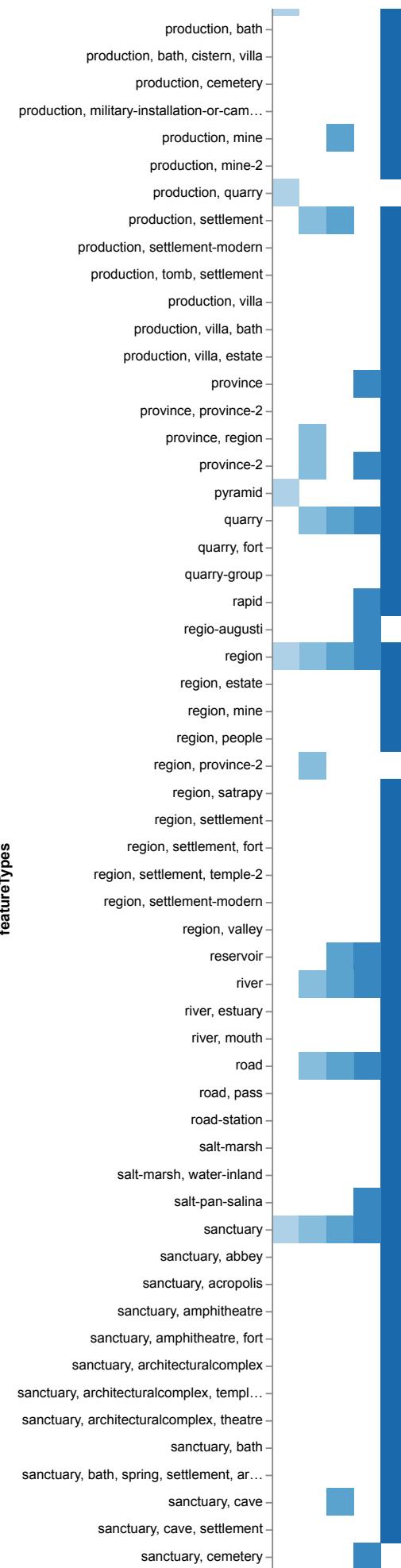
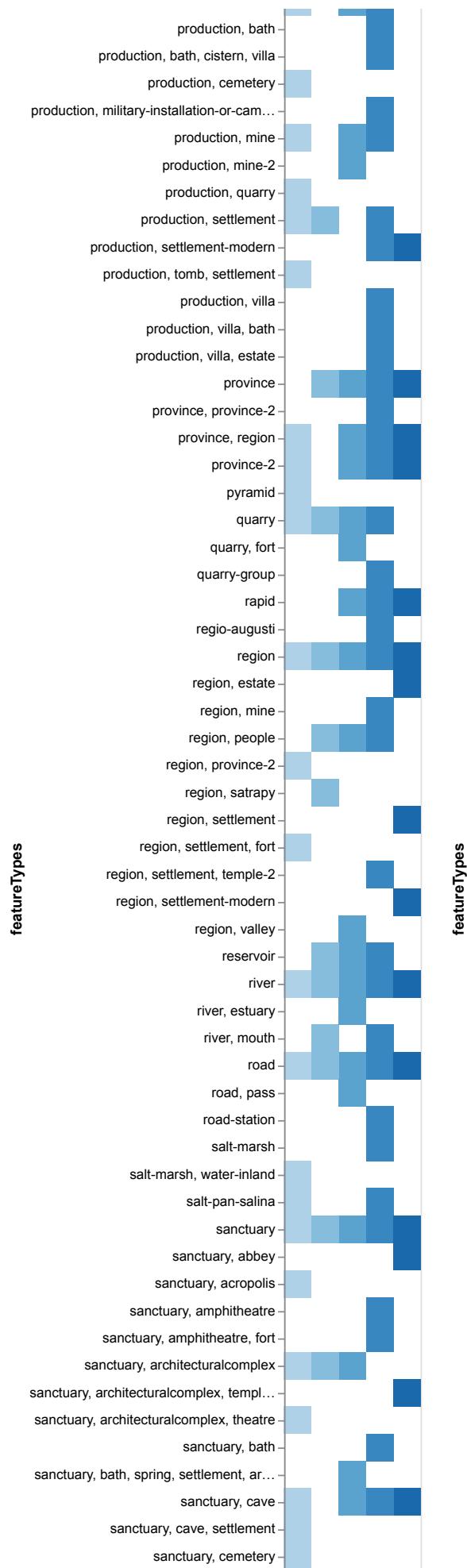




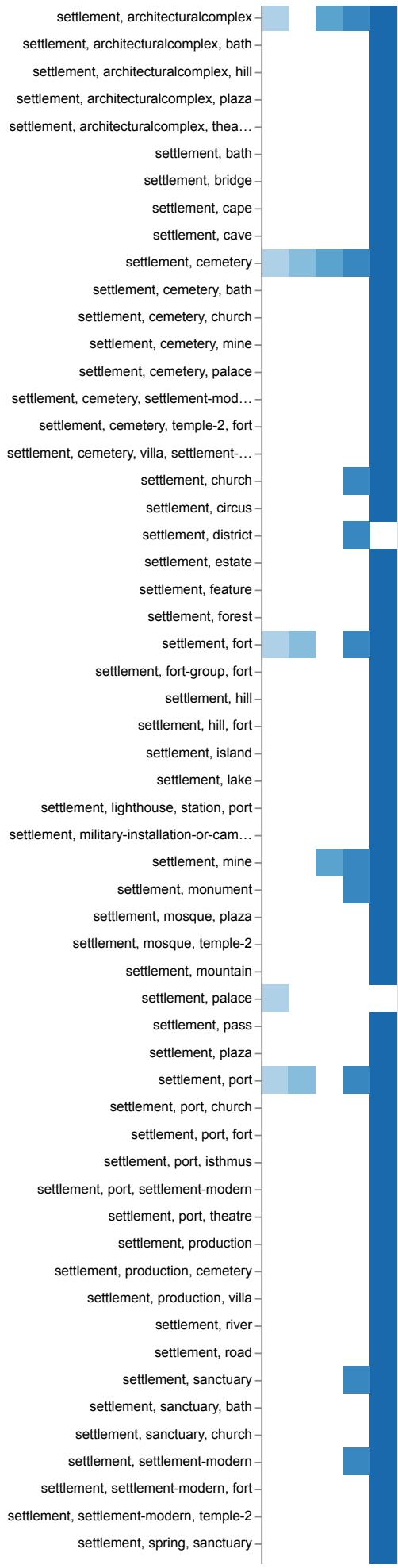
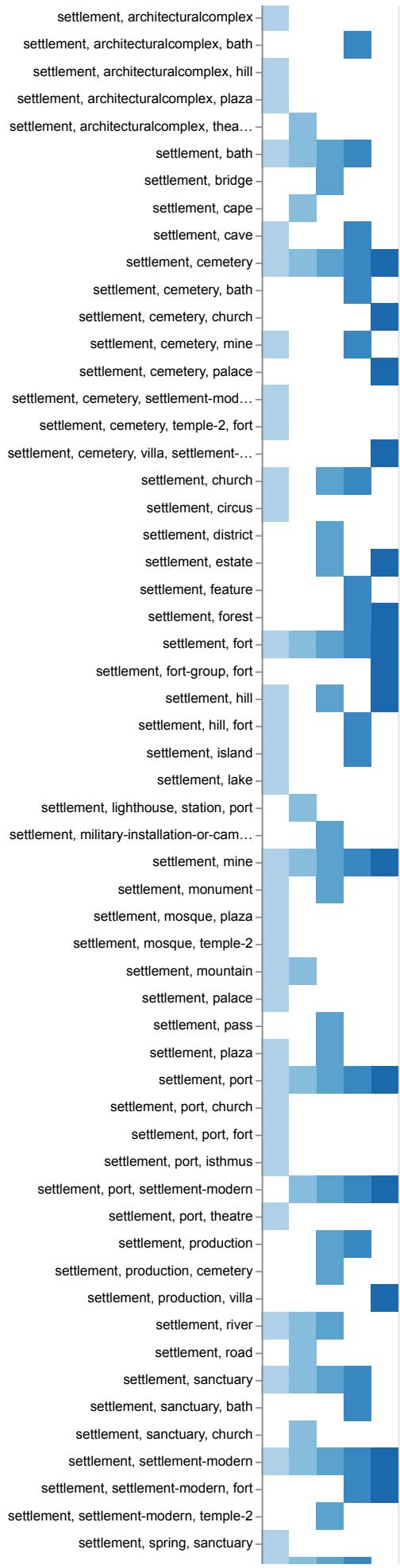


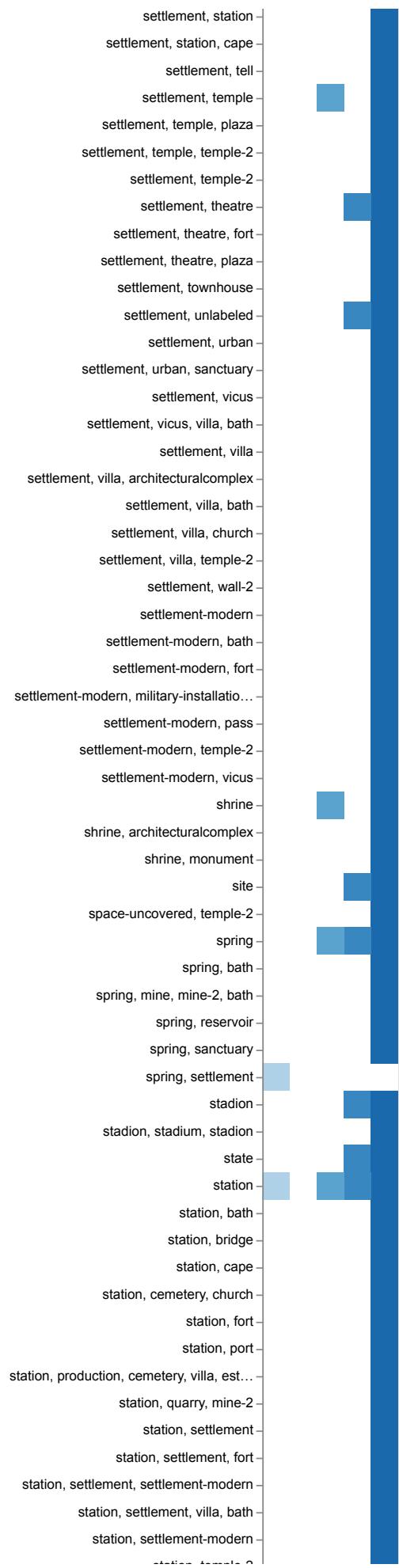
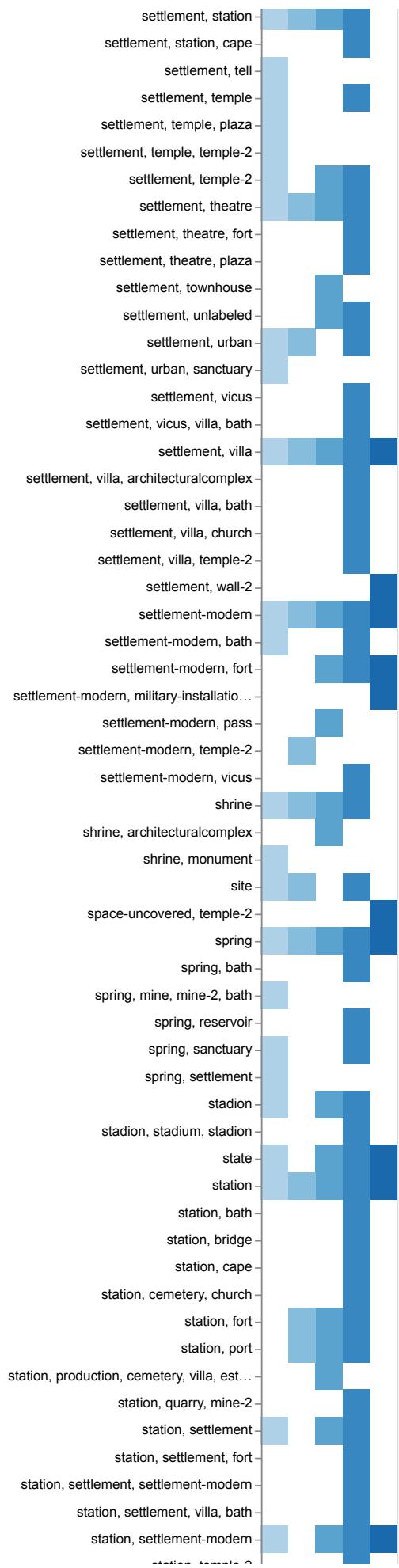


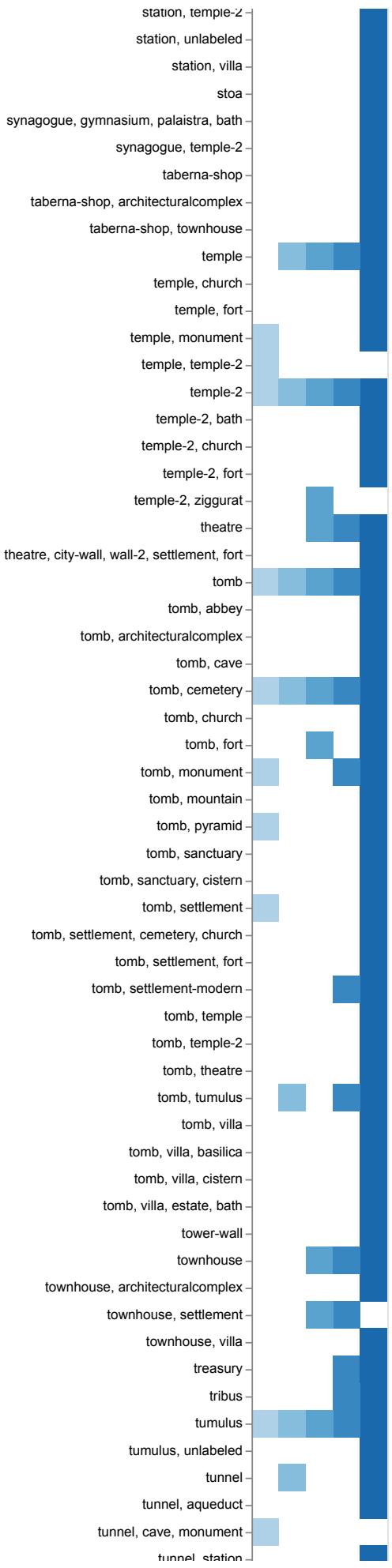
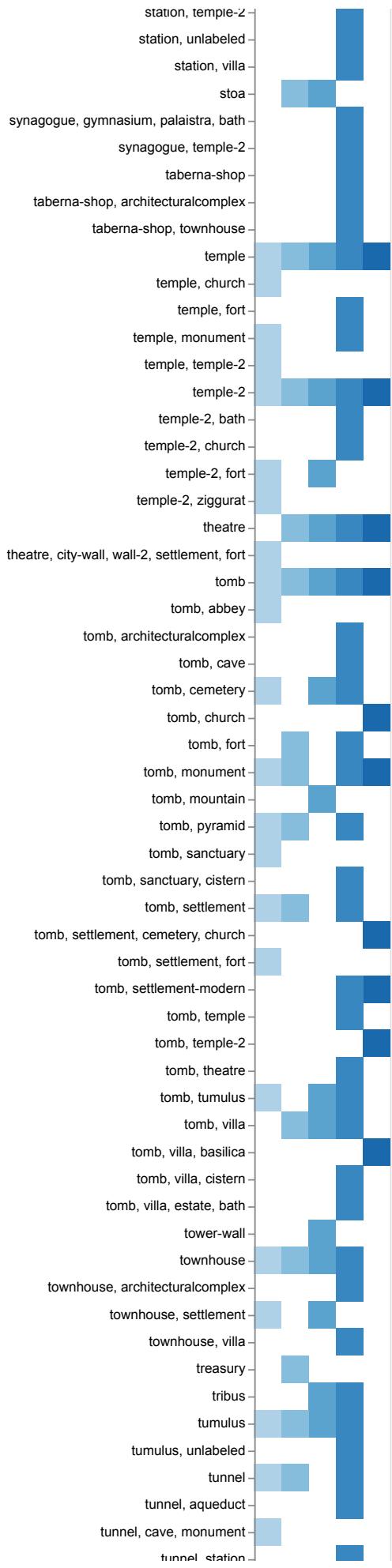


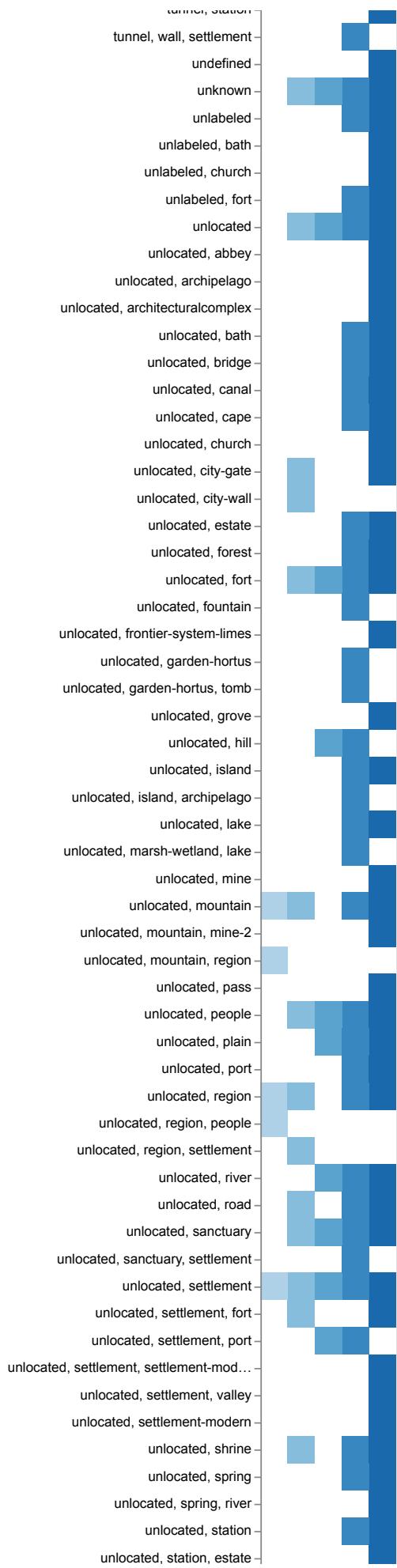
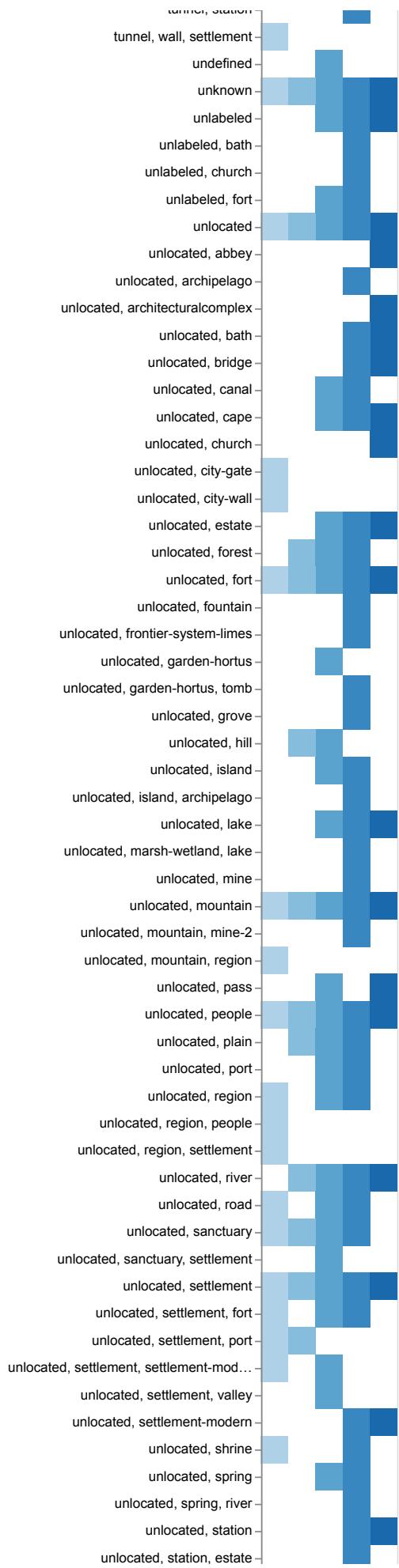


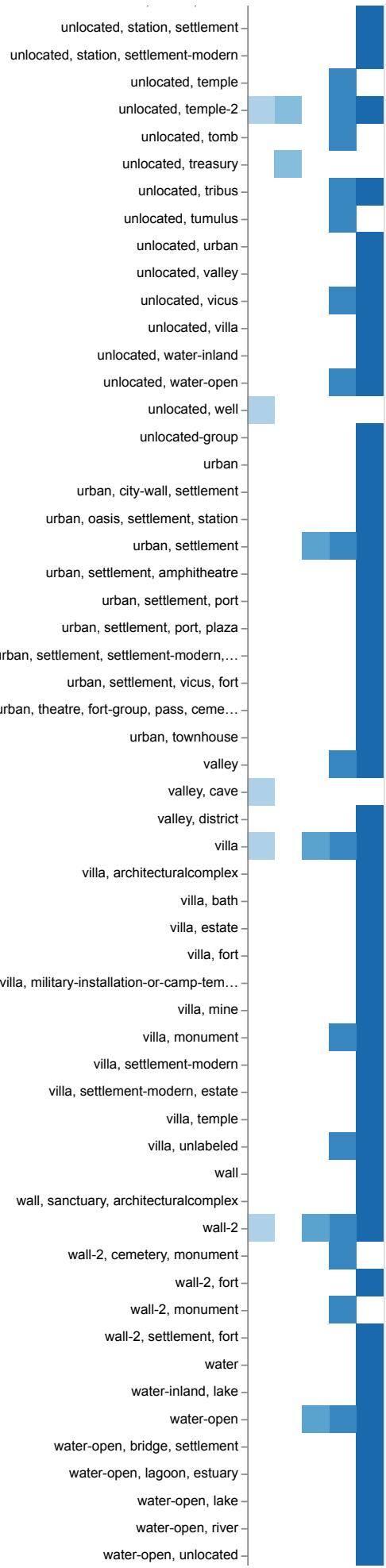
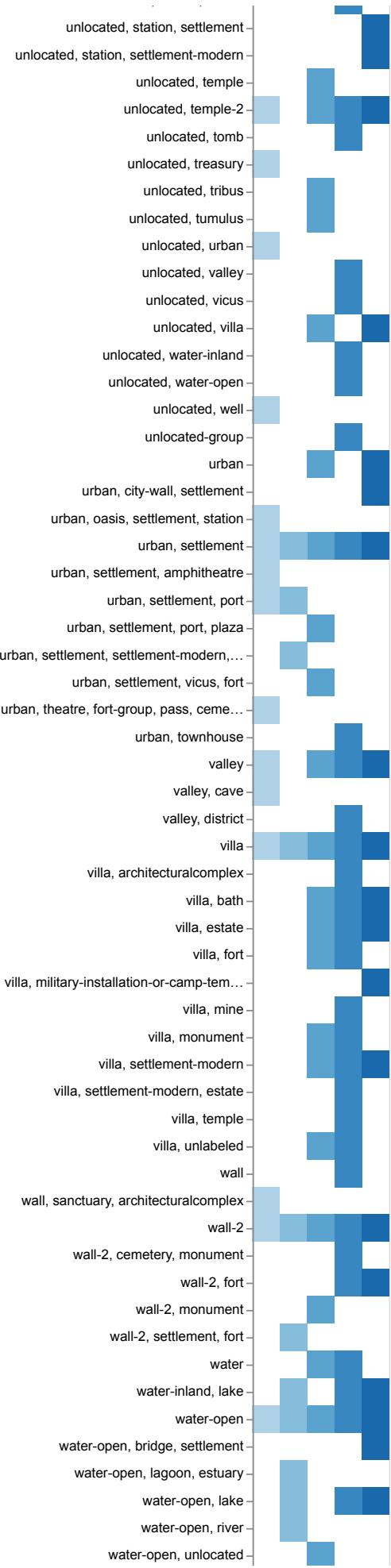


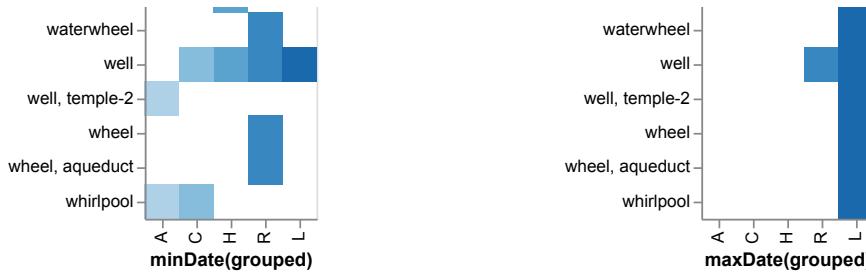












## What are all visual mappings used?

Given that date is minDate or maxDate

### *x position*

date (grouped and sorted according to the group)

### *y position*

featureTypes

### *color*

date (grouped and sorted according to the group)

### *tooltip*

count of records

## Was there any special data preparation done?

The data has been grouped to match the pleides README on `timePeriods` such that the records fall into the following bins "... 'A' (1000-550 BC), 'C' (550-330 BC), 'H' (330-30 BC), 'R' (AD 30-300), 'L' (AD 300-640)".

## What are the limitations of your design?

The visualization takes up to much space. This could be improved by using a smaller sample size or possibly, grouping the `featureTypes` into a smaller list of sub-categories such that its records are easier to see at a glance. Data processing to split the `featureTypes` column by the delimiter could also help.

# Visualization 5

## What can we learn from the visualization?

To understand how features that existed in every time period are distributed throughout those time periods

## What is the name for the type of visualization(s) used?

Stacked and normalized stacked bar charts

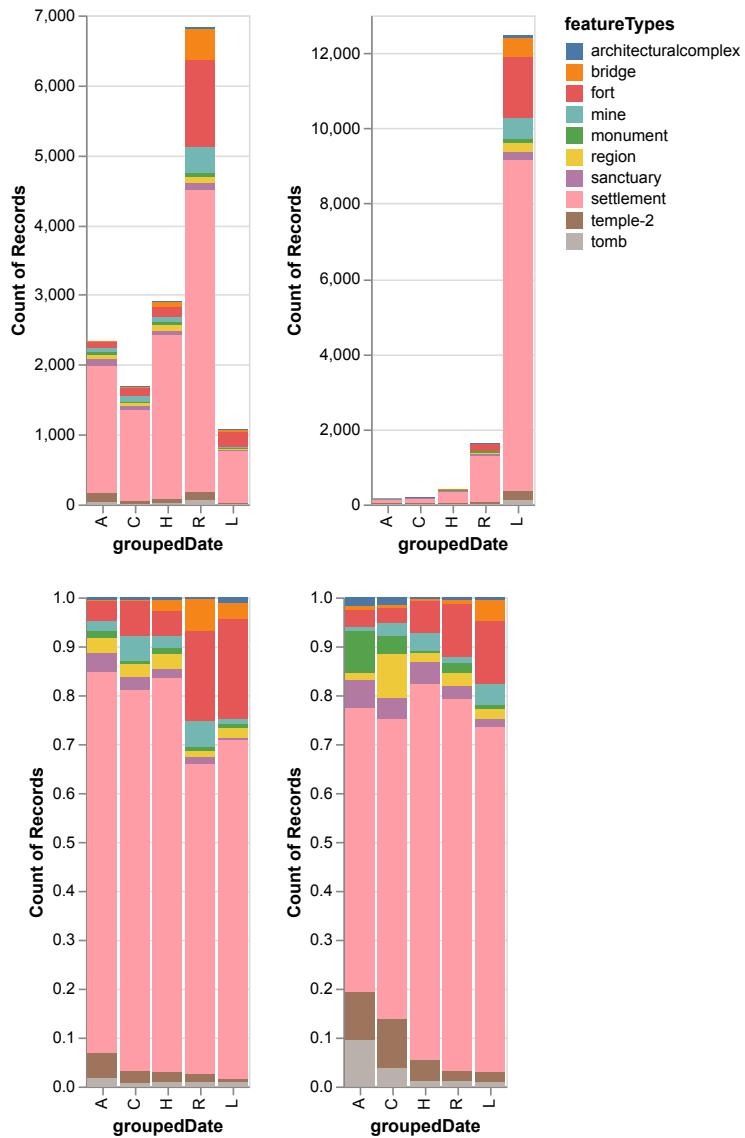
```
import altair as alt

places = ('https://raw.githubusercontent.com/SwanseaU-TTW/csc337_coursework1/master/ple'
popularFeatureTypes = ['settlement', 'architecturalcomplex', 'fort', 'temple-2', 'bridge',
                       'temple-1', 'gate', 'wall', 'excavation', 'residential']

def stackedBarChart(column):
    return (
        alt.Chart(places).transform_calculate(
            groupedDate=f'datum.{column} < -550 ? "A" : datum.{column} < -330 ? "C" : da
        ).mark_bar().encode(
            alt.X("groupedDate:0", sort=['A', 'C', 'H', 'R', 'L']),
            alt.Y('count()'),
            color='featureTypes:N',
            tooltip=['groupedDate:N', 'featureTypes:N', 'count()']
        ).transform_filter(
            alt.FieldOneOfPredicate(field='featureTypes', oneOf=popularFeatureTypes),
        ).interactive()
    )

    alt.vconcat(
        alt.hconcat(
            stackedBarChart('minDate'),
            stackedBarChart('maxDate')
        ),
        alt.hconcat(
            stackedBarChart('minDate').encode(alt.Y('count()', stack='normalize')),
            stackedBarChart('maxDate').encode(alt.Y('count()', stack='normalize'))
        )
    ).properties(
        title='Stacked bar charts and corresponding normalized stacked bar charts showing po
    ).configure_title(orient='top', anchor='middle')
```

Stacked bar charts and corresponding normalized stacked bar charts showing popular features' minDate and maxDate within each time period



## What are all visual mappings used?

Given that date is minDate or maxDate

### x position

date (grouped and sorted according to the group)

### y position

count of records

### color

featureTypes

### tooltip

date (grouped), featureTypes, count of each featureType

## **Was there any special data preparation done?**

The data has been grouped to match the pleaides README on `timePeriods` such that the records fall into the following bins "... 'A' (1000-550 BC), 'C' (550-330 BC), 'H' (330-30 BC), 'R' (AD 30-300), 'L' (AD 300-640)". The data has also been filtered such that only a subset of `featureTypes` that are known to exist across all `timePeriods` are observed.

## **What are the limitations of your design?**

This visualization could be improved by allowing users the ability to zoom into the chart. This visualization could be improved by randomly choosing what `featureTypes` to observe. Better still, improvements could be made by dynamically choosing what records to observe depending on various criteria, especially across the identified `timePeriods`. Further, input elements such as checkboxes, dropdowns and/or radio buttons could be used to allow dynamical selection of what criteria is used on the visualization.