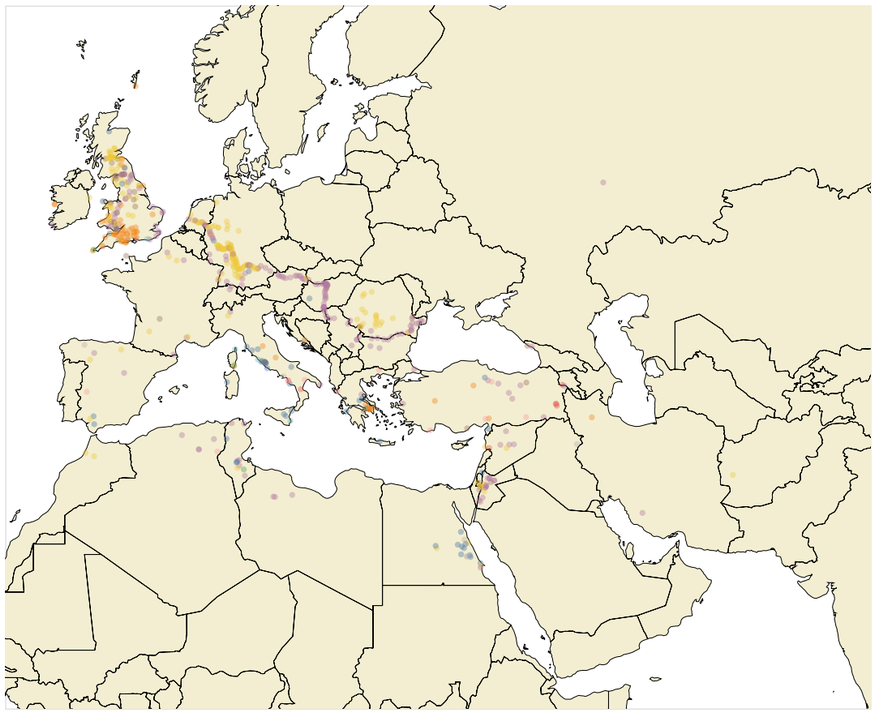
# Distribution of features around the world

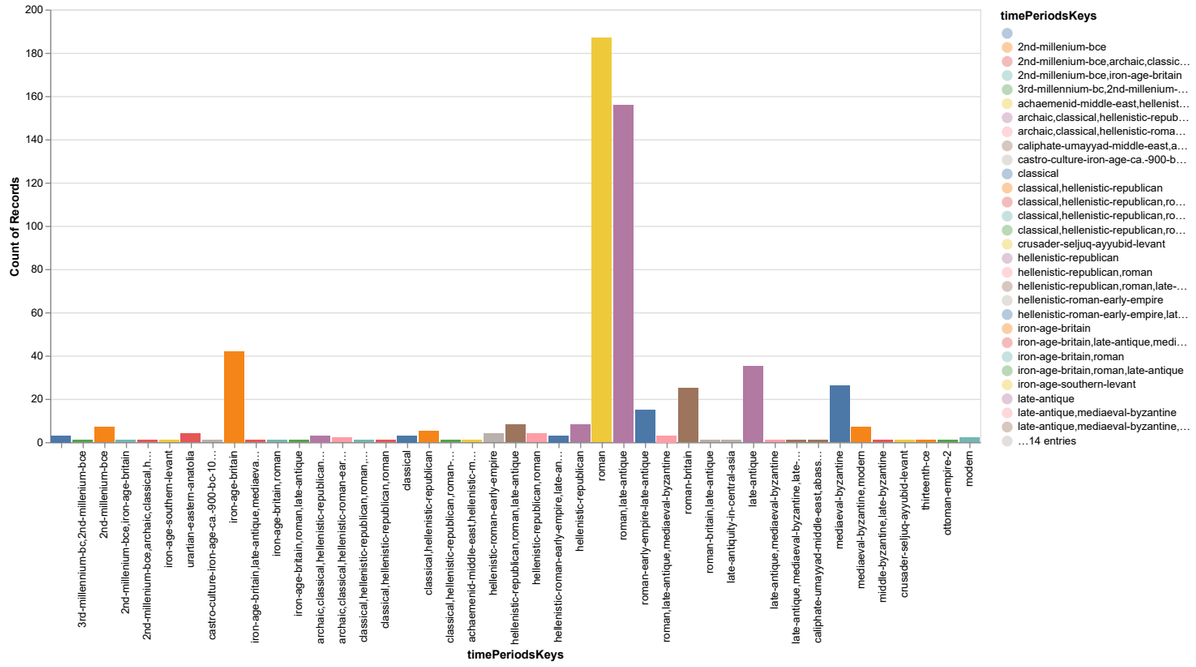
## Map scatterplot / bar chart

The aim of this visualization is to better understand the distribution of selected features throughout the world and their frequency in different time periods, as shown in the bar chart next to the map.

Note: The time periods in the bar chart are ordered by the 'mean' values of the minDates from the respective time periods.

Modifications and filters:
Unknown feature types are not displayed
Removed the commas from the features in the dataset






mySelection_featureType fort

## Observations

This visualization does not show many details unless you are looking for a specific feature type, so let's do that!

If we search for 'fort' we can see that we seem to have a line pattern in the middle of Europe between the Roman and the late-antique Roman periods. It would seem as if the Roman Empire had built even more forts between these time frames. Moreso, the positions of the forts start to make more sense when juxtaposed with the map of the late-antique period, as the line made by the forts marks the border between the Roman Empire and Germania Antiqua.

An interesting finding when searching for 'settlement' is how much the settlements have gradually spread out throughout the Roman, Late-Antique and Helenistic periods. We first see many settlements in the Roman Empire (in Italy), and then gradually spreading out with even greater numbers as the years go. The number of settlements does not just spread in Europe, it also spreads in Tunisia and the Middle-East during the late-antique period.

And, if we search for 'well', we will find a single sad well in Sardinia. While it would appear to be quite inconspicuous and lonely, it's actually one of the most famous tourist attractions in Sardinia, called 'Pozzo sacro di Santa Cristina'. If you ever find yourself going to Sardinia, give it a visit!

# Limitations

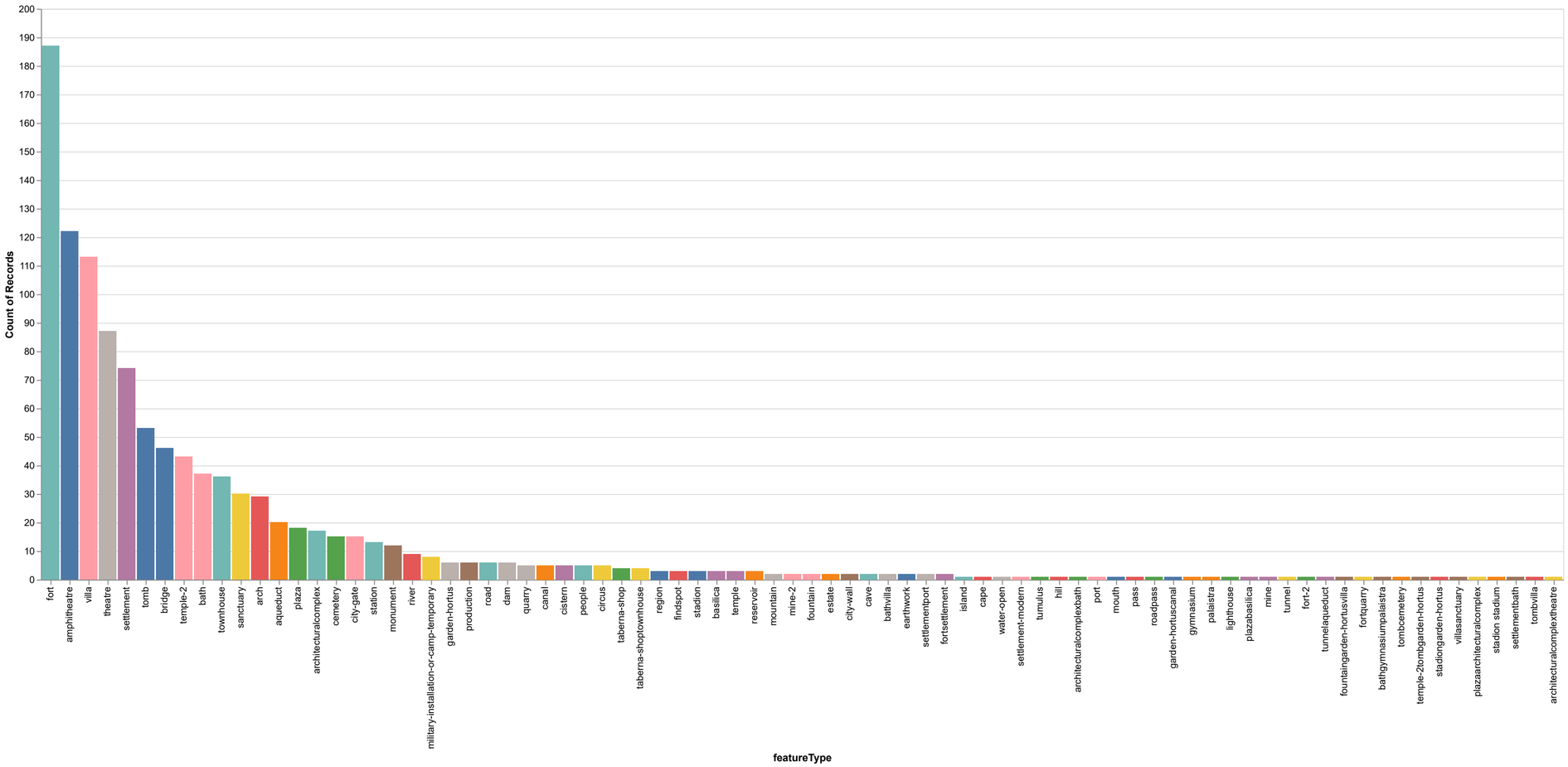
Unfortunately, this visualization is not perfect. There is no way of zooming in on the map, so certain patterns are not obvious. Moreover, if there are many features scattered throughout the world, the visualization will fail to give much insight.
The biggest limitation of this visualization is how the selection require knowledge of the features in the dataset. As such, it is impossible to explore without the original data at hand.
This visualization can be improved by adding a whole list of all of the different features and a zoom feature.

# Distribution of features around the world

## Single view bar chart histogram

This visualization illustrates the predominent feature types found in a given time period.

Modifications and filters:
Unknown and empty strings feature types are not displayed
Removed the commas from the features in the dataset




mySelection_timePeriodsKeys roman

## Observations

Off the start, without any filters, we can see that the most prominent features in the dataset are settlements, with over 2200 entries in the data, followed by almost 600 rivers and forts and other features.
The more we go to the right, we can find very niche structures with around a single entry in the data, such as 'villabath' or 'sanctuarysetlement'. These are combinations of prominent features found in the data that happened to be built together.

If we search for the 'roman' time period, we can see that they have quite the diverse set of structures, with the most prominent feature being the 'fort', which only makes sense given their supremacy during the respective period, a strong empire needed the best defense possible to protect itself. On top of this, they have plenty of amphitheatres and threatres. It is known that the world has culturally developed a lot during these times, but it is definitely interesting to observe just how popular these forms of entertainment were.

# Limitations

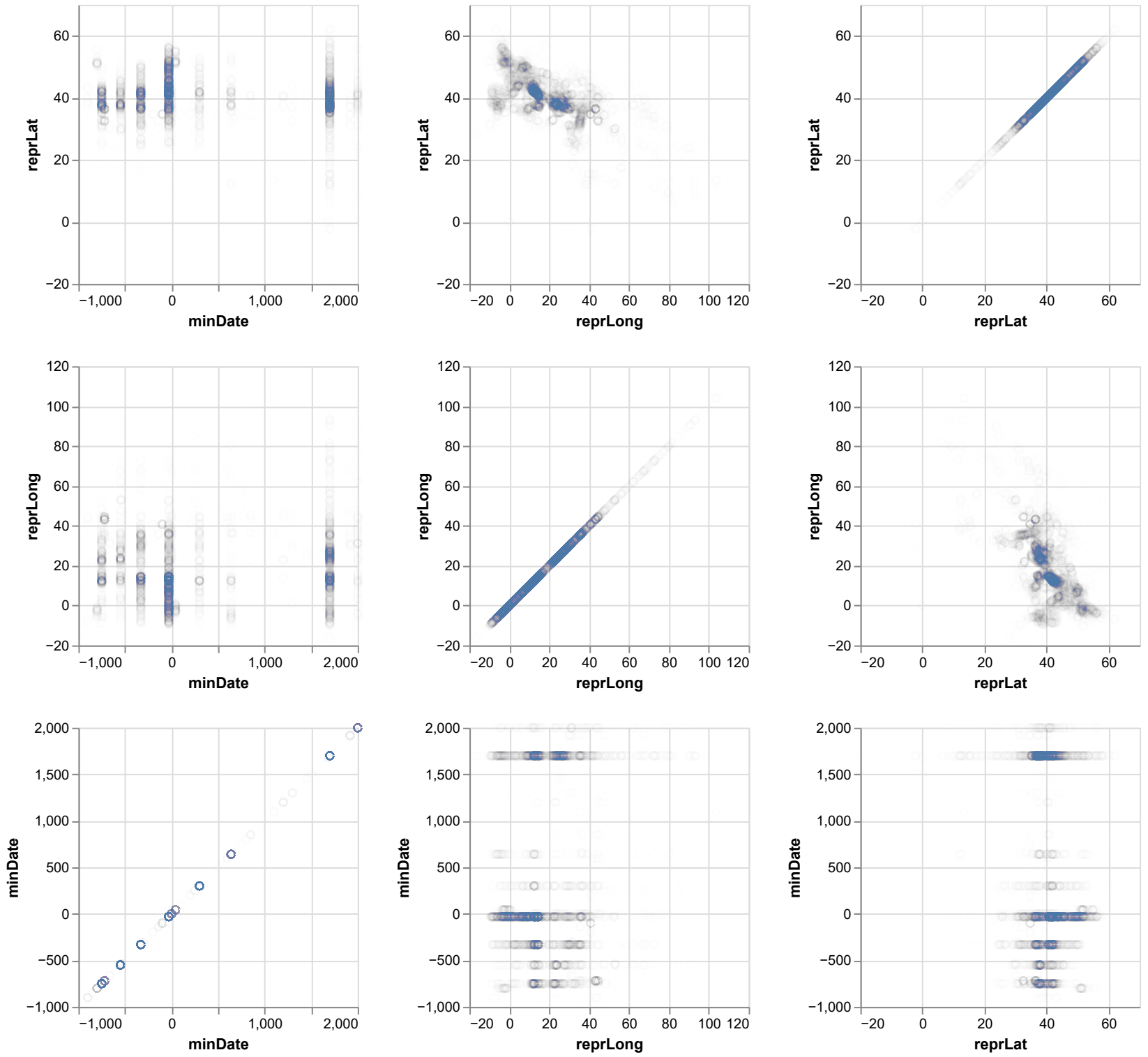
Much like the first visualization, this one is limited by the fact that it requires knowledge from the user about the different time periods. This can be fixed by adding a list of all of the time periods on the selection. On top of this, this visualization shows no geographical information about the features.

# Distribution of features around the world

## Scatter plot matrix

The aim of this visualization is to analyse the movement of structures throughout the years.

Modifications and filters:
Only displaying items from minDate 1000 BC onwards
Latitude and longitude values bigger than 999.99 are filtered
'Unknown' features are not displayed (efficiency purposes)




## Observations

If we look at the reprLong/minDate and reprLat/minDate graphs, we can observe how the majority of structures were built around the (41, 30) coordinates, while slowly the latitude starts to greatly spread in the range between 55-30 around the year 0 AD. In other words, if we were to visualize this on a mercator projection map, we would observe how structures started to spread vertically, from Northern Europe all the way to North Africa.

On the other hand the longitude seems to have decreased throughout the years, with the mean of the older structures from around 1000 BC being at a longitude of around 20, and with a mean latitude of 40 in that time, we can deduct these structures are mainly from the Roman Empire. However, the median longitude decreases to around 10, and the range is evenly spread between 40 and -10. This is most likely because of the spread of civilization throughout europe and the middle east.

# Limitations

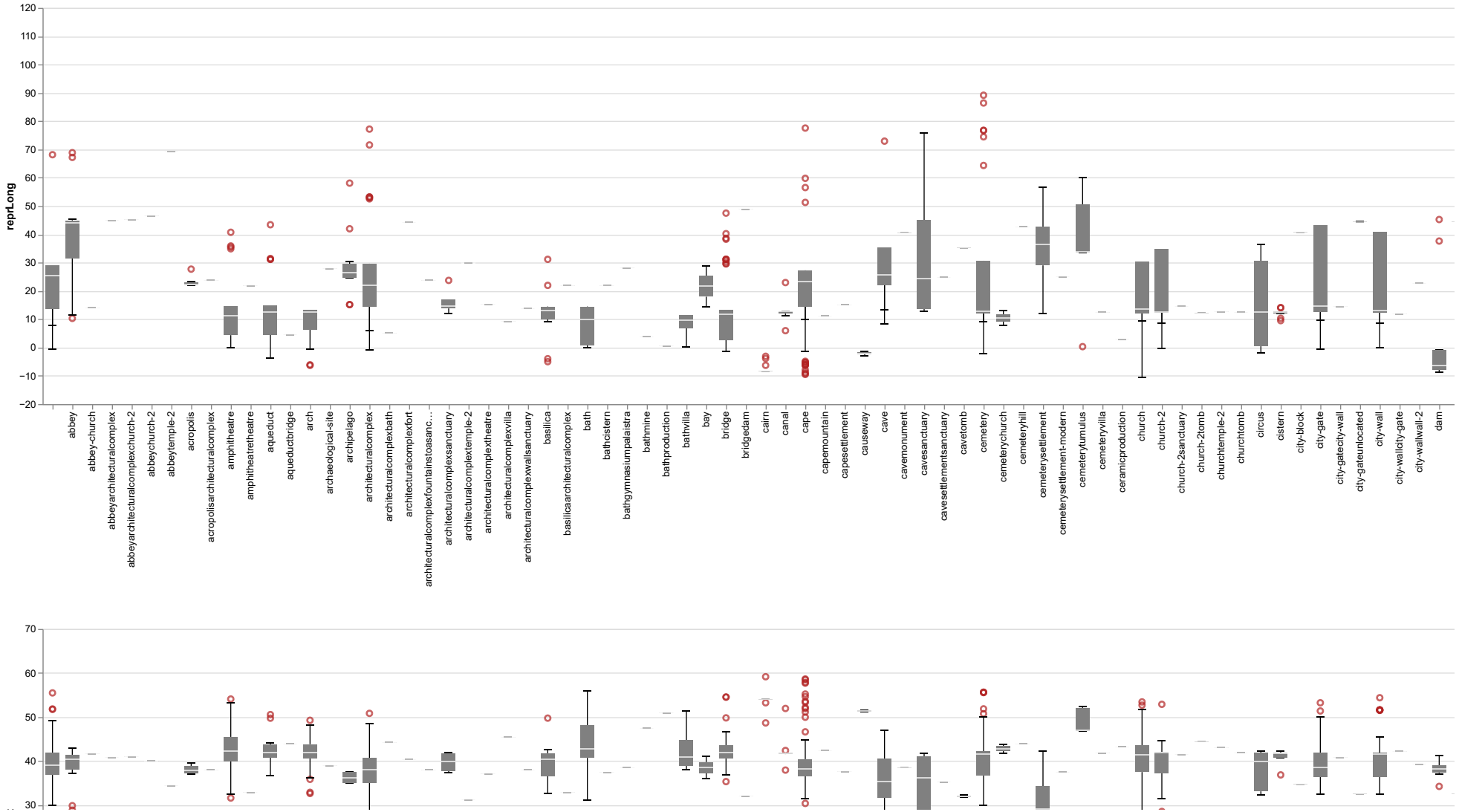
The biggest limitation of this visualization is the fact that it requires a map to properly visualize the movement and spread of society in a geographical context.

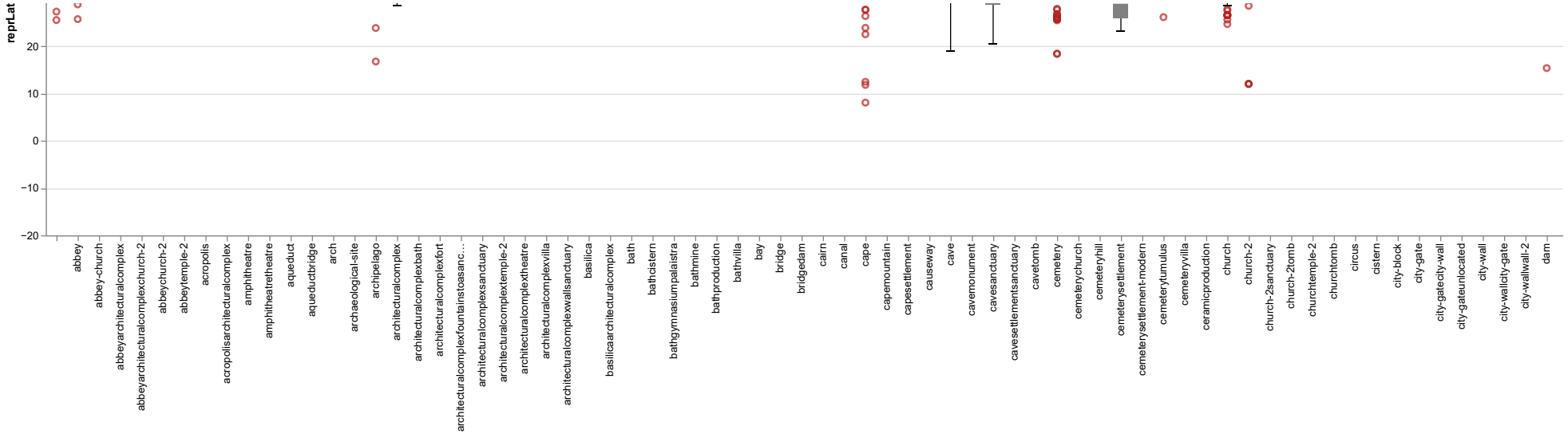
# Longitude and latitude variances of feature types

## Box plot

The aim of this visualization is to analyse the spread of feature types on the map and observing the outliers in each category.

Modifications and filters:
Removed the commas from the features in the dataset
Only displaying data with longitude and latitude smaller than 1000






## Observations

Looking at the graphs in general, we can see that most of the structures in the data tend to be around the latitude of 40 and a longitude of 20, which makes sense, given that the data primarily covers Europe and buildings from the Roman Empire.

However, we can identify some interesting outliers in the data. Most of these can be seen in the range 60-100 of the outliers, with these features mainly being from Asia. These features are primarily settlements or geographical features, such as rivers and islands.

# Limitations

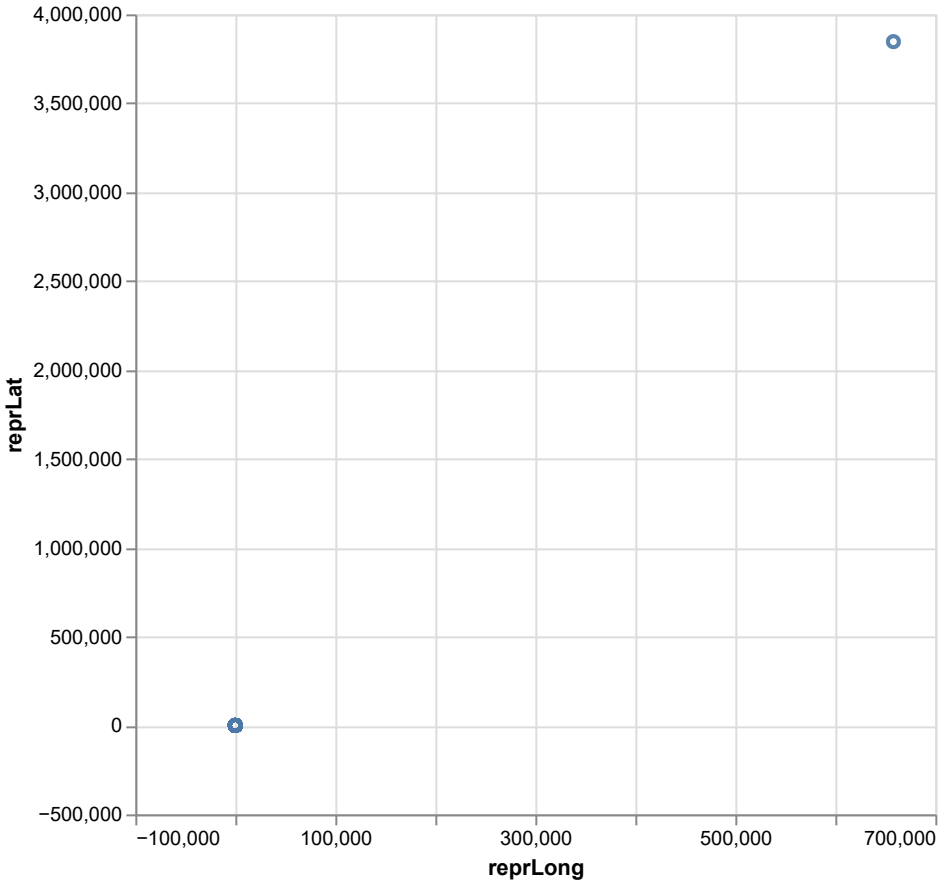
The primary limitation of this visualization is how features are not clearly linked with each other. So while we know we have outliers in the latitude graph, we have no way of telling what connection they have with the elements in the logitude graph.

# Analysing the outliers in the data

## Scatterplot

This is a simple unfiltered scatterplot of the data that uses the latitude as the Y axis, and the longitude as the X axis. However, it does not really look like it, given that we can only see two points on the map. This is because of two peculiar outliers in the data, more specifically two tower-walls at the coordinates of (3846187.763, 658013.3315) and (3846486.449, 658137.5669).Naturally, these values are invalid, so we can assume that the decimal points are in the wrong places, and instead, the values are (38.46187763, 65.80133315) and (3846486.449, 658137.5669), which yields two location in Uzbekistan.

Modifications:
None




# Limitations

The scatterplot shows us little to no information about the respective features, such as their periods or feature types.