

## Analysis1 - Luna Melek Gulec

### My primary research question(s)/interests for this project is:

"What is the relationship between earthquake magnitude and depth, how do these characteristics interact across different event types and reporting networks, and how measurement systems influence these interpretations?"

### This exploration focuses on:

- Quantitative Correlation: Examine the relationship between magnitude ( `mag` ) and depth ( `depth` ) for events, focusing mainly on earthquakes).
- Distributions: Analyze how magnitudes vary for shallow versus deeper events, and explore patterns in the engineered ordinal columns ( `mag_ordinal` and `depth_ordinal` ).
- Feature Relationships: Investigate the frequency of `mag_ordinal` and `depth_ordinal` categories ("minor", "moderate", etc. and "shallow", "deep", etc.) earthquakes across network sources ( `net` ) to identify regions with more major events and understand patterns. Understand how magnitude differs by network sources and measurement systems ( `magType` ).
- Comparative Statistics: Compare mean magnitude and depth for each magnitude category to understand how depth correlates with earthquake intensity.

## Dataset and Environment Setup

```
In [1]: import pandas as pd
import altair as alt
import numpy as np

In [2]: earthquakes = pd.read_csv('../data/processed/ordinal_data.csv')
```

## Exploratory Data Analysis (EDA)

### A. Magnitude vs. Depth

#### Distribution of Earthquake Magnitudes by Event Type

```
In [3]: # disable the "max rows = 5000" setting
alt.data_transformers.disable_max_rows()

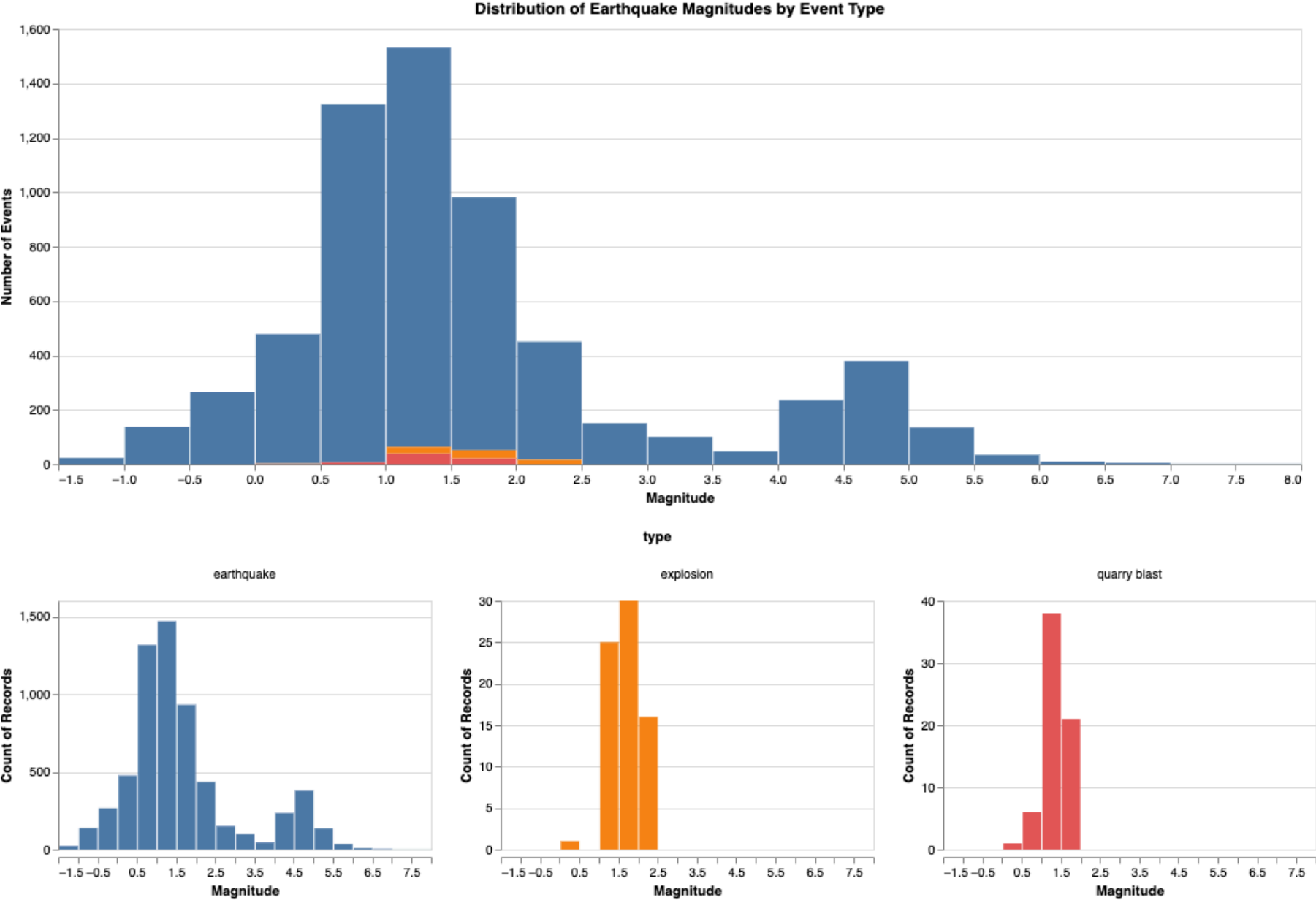
# to be able to view all charts in GitHub
alt.renderers.enable("mimetype")
```

```
# 1. layered histogram (all event types) (adapted from tutorial 5 IMDB ratings binned histogram)
layered_mag = alt.Chart(earthquakes).mark_bar().encode(
    alt.X('mag:Q', bin=alt.BinParams(maxbins=20), title='Magnitude'),
    alt.Y('count()', title='Number of Events'),
    alt.Color('type:N', title='Event Type'),
    tooltip=['count()', 'type']
).properties(width=1000, height=350, title='Distribution of Earthquake Magnitudes by Event Type')

# 2. faceted chart per event type (separate chart object)
faceted_mag = alt.Chart(earthquakes).mark_bar().encode(
    alt.X('mag:Q', bin=alt.BinParams(maxbins=20), title='Magnitude'),
    alt.Y('count()'),
    alt.Color('type:N', legend=None),
    alt.Column('type:N'),
    tooltip=['count()']
).properties(width=300, height=200).resolve_scale(y='independent')

# 3. combine
mag_viz = layered_mag & faceted_mag
mag_viz
```

Out[3]:



1. Top = layered histogram
- This layered histogram explores the distribution of earthquake magnitudes.
  - There is one dominant peak with a magnitude of around 1.5, the magnitude of almost 1,600 events.

- It also reveals a slight right tail with a secondary, smaller peak near magnitude 5.0, though with many less events of around 4.0.

2. Bottom = faceted histograms by event type

- To take this further, faceting by event type with independent scales details the distribution per event type. It shows that earthquakes and quarry blasts peak with a magnitude of around 1.5, whereas explosions peak slightly higher near 2.0.

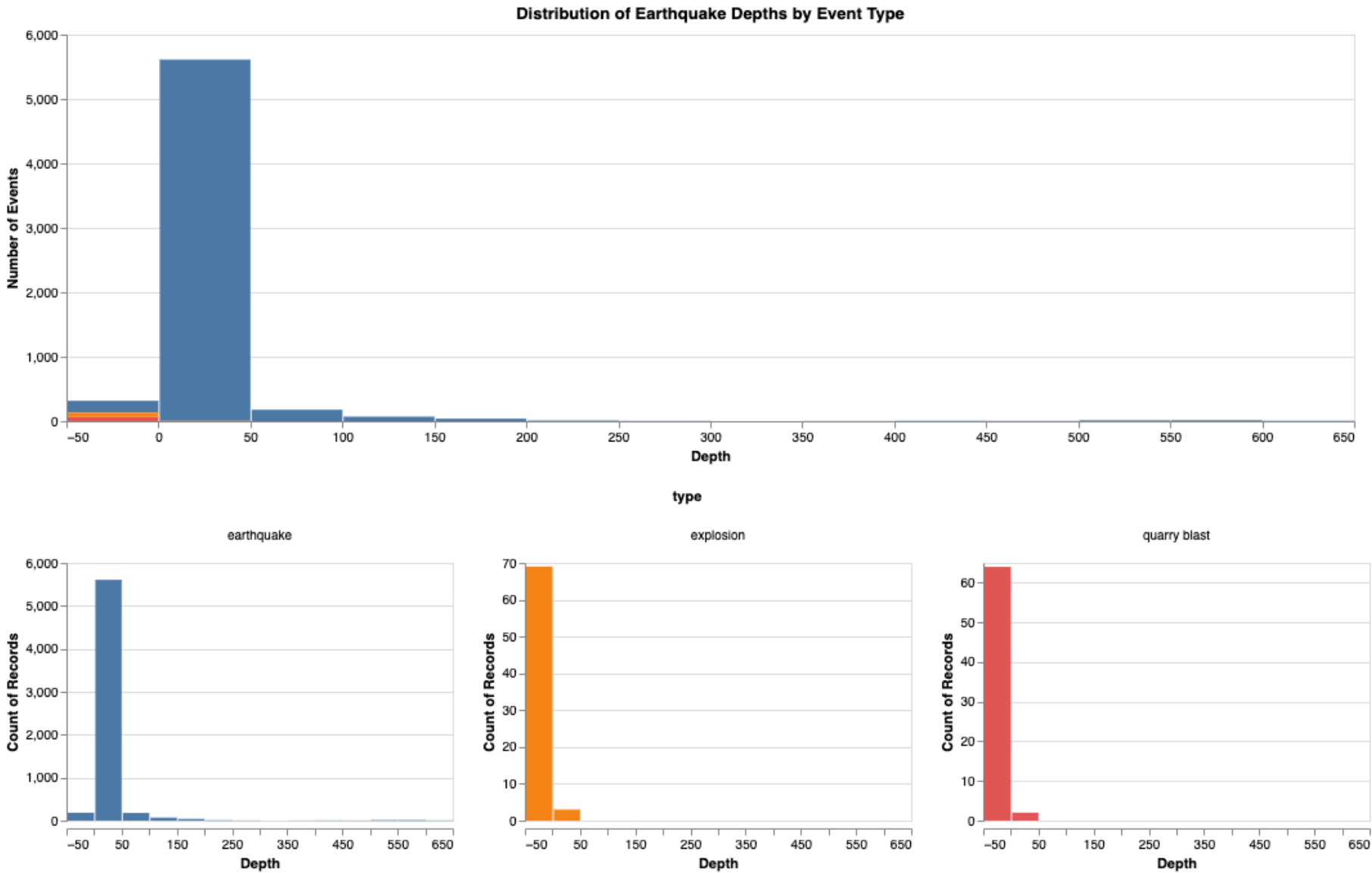
## Distribution of Earthquake Depths by Event Type

```
In [4]: # 1. layered histogram (all event types) (adapted from tutorial 5 IMDB ratings binned histogram)
layered_depth = alt.Chart(earthquakes).mark_bar().encode(
    alt.X('depth:Q', bin=alt.BinParams(maxbins=20), title='Depth'),
    alt.Y('count()', title='Number of Events'),
    alt.Color('type:N', title='Event Type'),
    tooltip=['count()', 'type']
).properties(width=1000, height=300, title='Distribution of Earthquake Depths by Event Type')

# 2. faceted chart per event type (separate chart object)
faceted_depth = alt.Chart(earthquakes).mark_bar().encode(
    alt.X('depth:Q', bin=alt.BinParams(maxbins=20), title='Depth'),
    alt.Y('count()'),
    alt.Color('type:N', legend=None),
    alt.Column('type:N'),
    tooltip=['count()']
).properties(width=300, height=200).resolve_scale(y='independent')

# 3. combine
depth_viz = layered_depth & faceted_depth
depth_viz
```

Out [4] :



1. Top = layered histogram

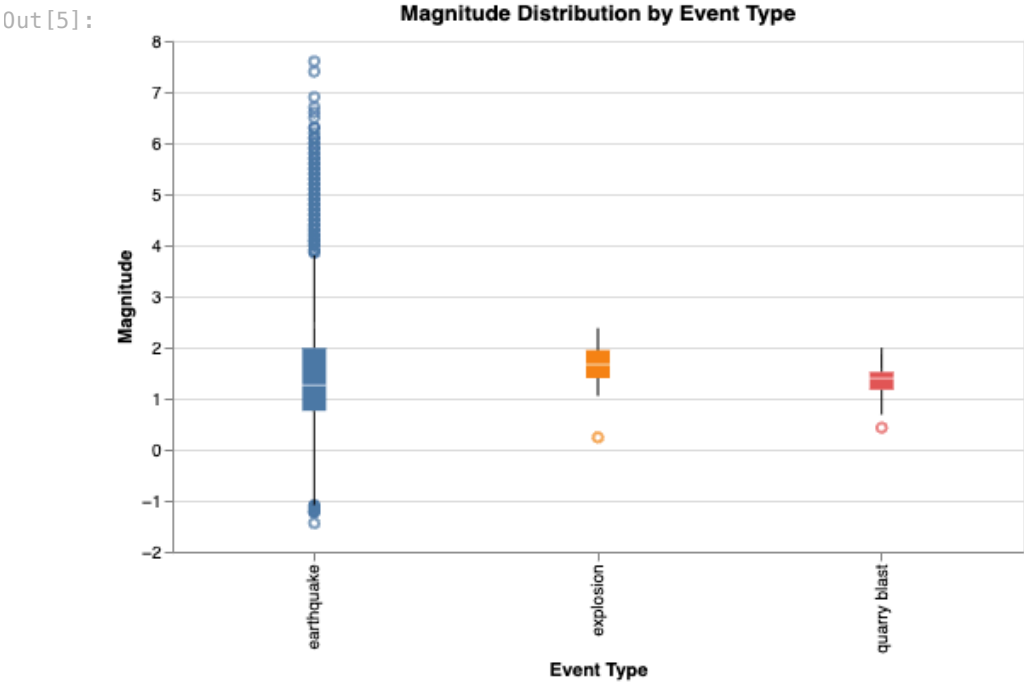
- This is the same histogram code but for depth instead.
- There is a clear mode of around 5,600 events at 0 to 50 depth , with under 500 events for other depths.
- After depth = 50, the number of events with increasing depths decreases steadily until depth = 250. However, there are a couple events (almost 0/negligible) of earthquakes with depths of 500 to 600.

2. Bottom = faceted histograms by type

- Again, faceting by event type with independent scales allows for viewing each event type distribution separately, as earthquakes have more records.
- This shows us how earthquakes have a mode of 0 to 50 depth (about 5,600 events), whereas both explosions and quarry blasts have a mode of depths between -50 and 50.

Magnitude Distribution by Event Type

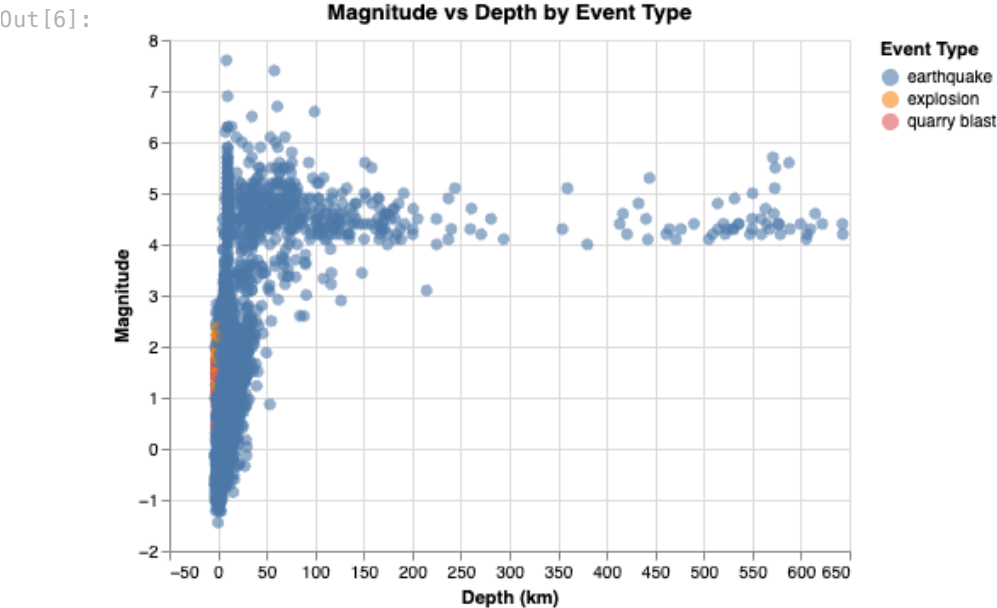
```
In [5]: alt.Chart(earthquakes).mark_boxplot(extent=1.5).encode(  
  x=alt.X('type:N', title='Event Type'),  
  y=alt.Y('mag:Q', title='Magnitude'),  
  color=alt.Color('type:N', legend=None),  
  tooltip=['mag', 'mag_ordinal']  
)  
.properties(width=500, height=300, title='Magnitude Distribution by Event Type')
```



- This boxplot highlights how:
  - Explosions have slightly higher average magnitudes.
  - Both explosions and quarry blasts have smaller ranges.
  - Earthquakes display a wider spread and many outliers, specifically clustered at higher magnitudes, suggesting that most earthquake events have lower (under 4) magnitudes, which is consistent with the earlier magnitude distribution.

Magnitude vs. Depth by Event Type - individual points

```
In [6]: alt.Chart(earthquakes).mark_circle(opacity=0.6, size=50).encode(
  x=alt.X('depth:Q', title='Depth (km)'),
  y=alt.Y('mag:Q', title='Magnitude'),
  color=alt.Color('type:N', title='Event Type'),
  tooltip=['type', 'depth', 'mag']
).properties(width=400, height=300, title='Magnitude vs Depth by Event Type')
```



- This scatter plot aims to similarly uncover the relationship between magnitude and depth, again coloured by event type, but viewing exact points.
- Focusing on earthquakes, as explosions and quarry blasts are few, depth typically stays under 200, with the majority of points clustering between depths 0-100, with varying magnitudes. However, deeper events (up to 650 km) corresponds to magnitudes of only 4-6.
  - Contrary to our initial expectations, the relationship between magnitude and depth is not strongly positive - i.e. deeper earthquakes are not necessarily stronger.
  - This could be due to the fact that most records have a magnitude of around 4-5.
  - This may also result from different tectonic regions or from detection bias (shallow quakes being more frequently recorded).

```
In [7]: # 1. MAG CHART
# order the categories of mag_ordinal
ordered_mag_cat = [
  'Minor (<4.0)',
  'Light (4.0-4.9)',
  'Moderate (5.0-5.9)',
  'Strong (6.0-6.9)',
```

```

    'Major (>=7.0)'
]

earthquakes['mag_ordinal'] = pd.Categorical(
    earthquakes['mag_ordinal'],
    categories=ordered_mag_cat,
    ordered=True
)

mag_chart = alt.Chart(earthquakes).mark_circle(opacity=0.5).encode(
    alt.X('depth:Q', title='Depth (km)'),
    alt.Y('mag:Q', title='Magnitude'),
    color=alt.Color(
        'mag_ordinal:O',
        scale=alt.Scale(domain=ordered_mag_cat, scheme='viridis'),
        legend=alt.Legend(title='Magnitude Category')
    ),
    tooltip=['mag', 'depth', 'mag_ordinal']
).properties(width=330, height=230, title='Magnitude vs Depth by Magnitude Category')

# 2. DEPTH CHART
# order the categories of depth_ordinal
ordered_depth_cat = [
    'Negative (<0 km)',
    'Shallow (0-70 km)',
    'Intermediate (70-300 km)',
    'Deep (>300 km)'
]

earthquakes['depth_ordinal'] = pd.Categorical(
    earthquakes['depth_ordinal'],
    categories=ordered_depth_cat,
    ordered=True
)

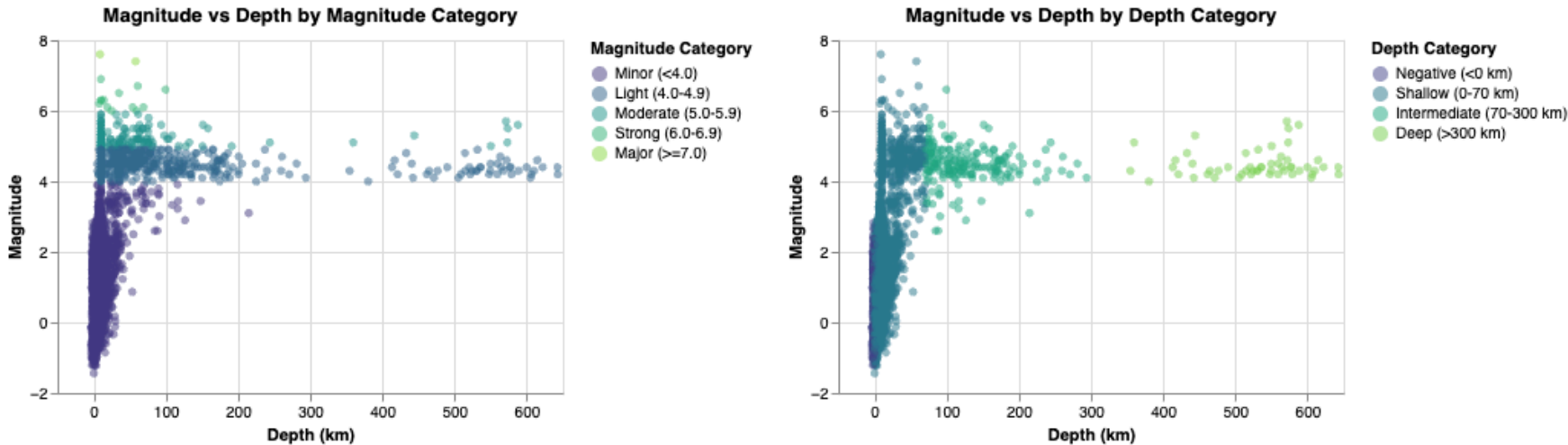
depth_chart = alt.Chart(earthquakes).mark_circle(opacity=0.5).encode(
    alt.X('depth:Q', title='Depth (km)'),
    alt.Y('mag:Q', title='Magnitude'),
    color=alt.Color(
        'depth_ordinal:O',
        scale=alt.Scale(domain=ordered_depth_cat, scheme='viridis'),
        legend=alt.Legend(title='Depth Category')
    ),
    tooltip=['mag', 'depth', 'depth_ordinal']
).properties(width=330, height=230, title='Magnitude vs Depth by Depth Category')

```



```
# VIEW TOGETHER
(mag_chart | depth_chart).resolve_scale(color='independent').resolve_legend(color='independent')
```

Out[7]:



- Colouring the same depth and magnitude scatter plot by magnitude category ( `mag_ordinal` ) provides a clearer distribution, and we can see clearly that the majority of earthquakes are minor (magnitude category under 4).
- Doing the same colouring by depth category ( `depth_ordinal` ) helps identify shallow vs. deep zones visually.
- The charts show that the majority of earthquakes have a shallow depth (0-70 km), with few as negative, and roughly equal amounts of intermediate (70-300 km) and deep (over 300 km), as well as magnitude clustered around 4.

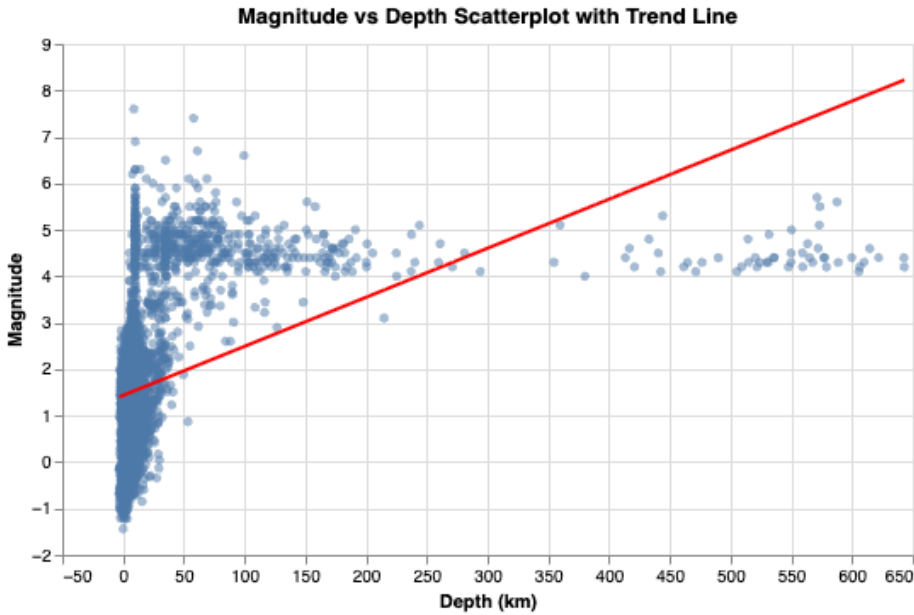
Understanding the correlation between Magnitude and Depth

```
In [8]: # adapted from tutorial 6 correlation scatter plot + trend line
scatter = alt.Chart(earthquakes).mark_circle(opacity=0.5).encode(
    alt.X('depth:Q', title='Depth (km)'),
    alt.Y('mag:Q', title='Magnitude'),
    tooltip=['mag', 'depth', 'type']
).interactive()

trend = scatter.transform_regression('depth', 'mag').mark_line(color='red')

(scatter + trend).properties(width=500, height=300, title='Magnitude vs Depth Scatterplot with Trend Line')
```

Out [8]:



```
In [9]: correlation_coef = earthquakes['depth'].corr(earthquakes['mag'])
print(f"Correlation coefficient: {correlation_coef:.3f}")
```

Correlation coefficient: 0.397

- The correlation between depth and magnitude is 0.397, which is weakly positive.
- While deeper earthquakes can be stronger on average, magnitude tends to depends on other geological factors (e.g. fault structure, tectonic stress).

B. Magnitude & Depth Groupby Summary Tables (Quantitative Analysis)

```
In [10]: earthquakes.groupby('type')[['mag', 'depth']].agg(['mean', 'std', 'count', 'max', 'min'])
```

Out [10]:

	mag					depth				
	mean	std	count	max	min	mean	std	count	max	min
type										
earthquake	1.615540	1.429820	6143	7.60	-1.44	17.221997	53.756168	6143	643.06	-3.18
explosion	1.653056	0.386697	72	2.37	0.24	-0.517778	0.954797	72	4.54	-1.86
quarry blast	1.360000	0.293158	66	1.99	0.43	-1.047121	1.276247	66	6.42	-2.00

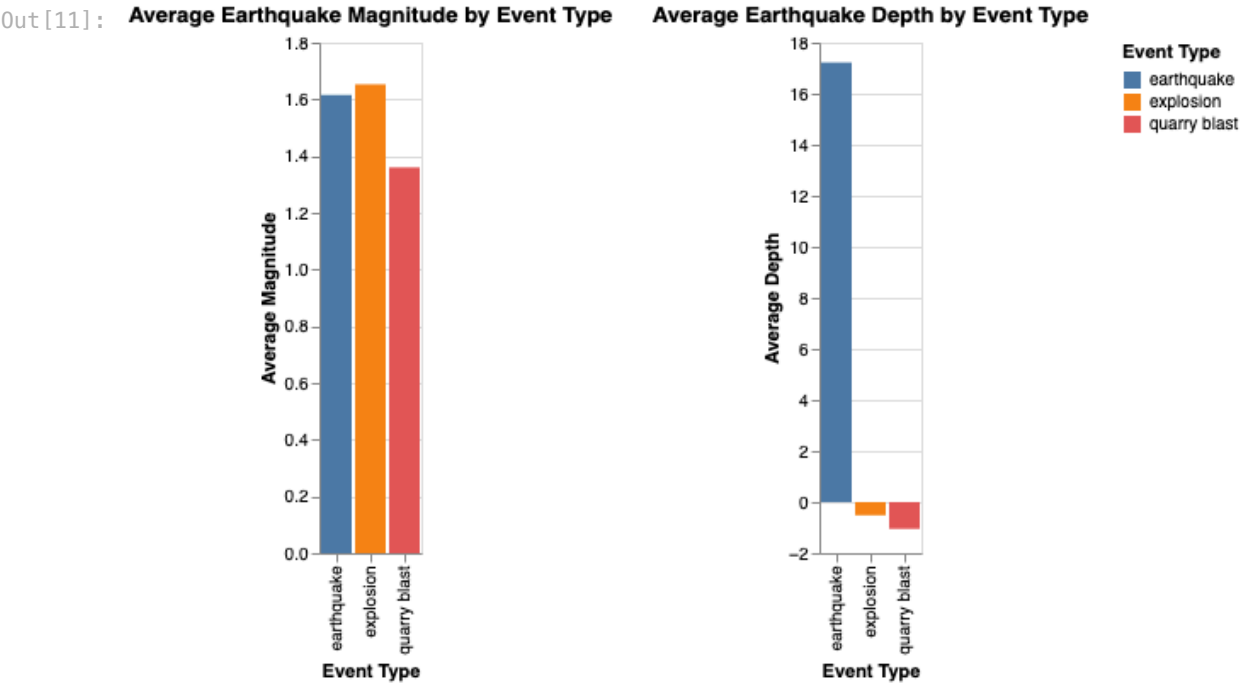
- The table shows comparable average magnitudes among event type but a stark difference in depths - earthquakes average ~17 km depth, compared to ~0 km for explosions and quarry blasts, with a standard deviation of 54 compared to 1. Depth also has a much wider range for earthquakes.
- Therefore, this indicates that depth varies quite a bit, and we can explore this further.
- To view magnitudes and depths of event types further:

```
In [11]: # view as bar chart
avg_by_type = earthquakes.groupby('type', as_index=False)[['mag', 'depth']].mean()

avg_mag = alt.Chart(avg_by_type).mark_bar().encode(
    alt.X('type:N', title='Event Type'),
    alt.Y('mag:Q', title='Average Magnitude'),
    alt.Color('type:N', title='Event Type'),
).properties(title='Average Earthquake Magnitude by Event Type')

avg_depth = alt.Chart(avg_by_type).mark_bar().encode(
    alt.X('type:N', title='Event Type'),
    alt.Y('depth:Q', title='Average Depth'),
    alt.Color('type:N', title='Event Type')
).properties(title='Average Earthquake Depth by Event Type')

avg_mag | avg_depth
```



- From the table and visualising the figures on bar charts, we can see that the average magnitude is pretty consistent among event types, whereas average depth shows clear differences between earthquakes (much higher) and both explosions and quarry blasts.

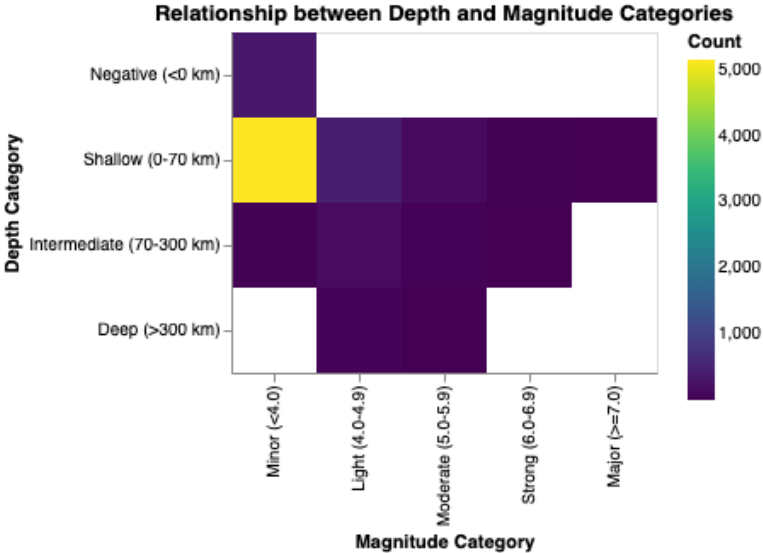
C. Magnitude & Depth Categories (Ordinal Analysis)

Comparing `depth_ordinal` and `mag_ordinal`

```
In [12]: agg = earthquakes.groupby(['depth_ordinal', 'mag_ordinal'], as_index=False, observed=True).size()

alt.Chart(agg).mark_rect().encode(
    alt.X('mag_ordinal:O', title='Magnitude Category', sort=ordered_mag_cat),
    alt.Y('depth_ordinal:O', title='Depth Category', sort=ordered_depth_cat),
    alt.Color('size:Q', title='Count', scale=alt.Scale(scheme='viridis')),
    tooltip=['depth_ordinal', 'mag_ordinal', 'size']
).properties(title='Relationship between Depth and Magnitude Categories', width=250, height=200)
```

Out[12]:

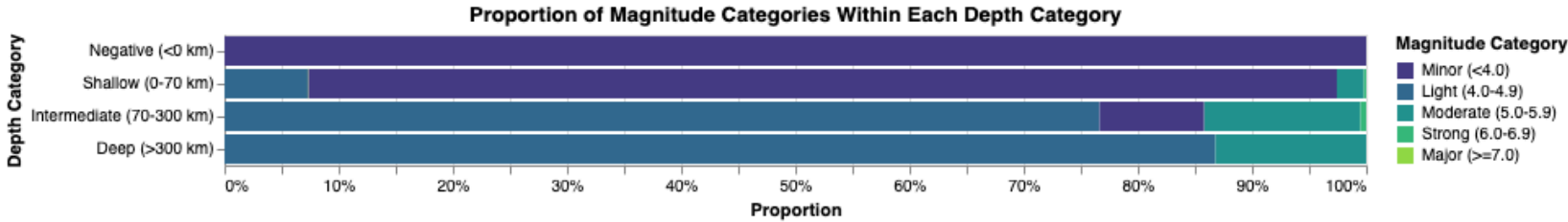


- This heatmap groups the data by `mag_ordinal` and `depth_ordinal`, and reveals that most events occur in shallow zones (0-70 km) and are classified as Minor (magnitude < 4.0).
- As depth increases, the proportion of Moderate and Strong magnitudes slightly rises, though data volume decreases with greater depths.
- It is also interesting to note that all Major earthquakes (magnitude >= 7.0) are shallow, which contradicts the initial assumption that major earthquakes must be deep and intense.

```
In [13]: alt.Chart(earthquakes).mark_bar().encode(
    alt.Y('depth_ordinal:N', title='Depth Category', sort=ordered_depth_cat),
    alt.X('count():Q', stack='normalize', title='Proportion'),
```

```
alt.Color('mag_ordinal:N', title='Magnitude Category', scale=alt.Scale(domain=ordered_mag_cat, scheme='viridis')),
tooltip=['depth_ordinal', 'mag_ordinal', 'count()']
).properties(title='Proportion of Magnitude Categories Within Each Depth Category', width=700)
```

Out [13]:



- This normalized bar chart is another way to view the relationship between the two ordinal categories - the proportion (%) of depth categories coloured by magnitude categories

D. Frequency of mag\_ordinal or depth\_ordinal by top 5 networks

```
In [14]: top_nets = earthquakes['net'].value_counts().nlargest(5).index
top_nets
```

Out[14]: Index(['nc', 'ci', 'tx', 'us', 'av'], dtype='object', name='net')

- The top 5 largest networks are nc (Northern California), ci (Southern California), tx (Texas), us (USGS National Earthquake Information Center), av (Alaska Volcano Observatory)

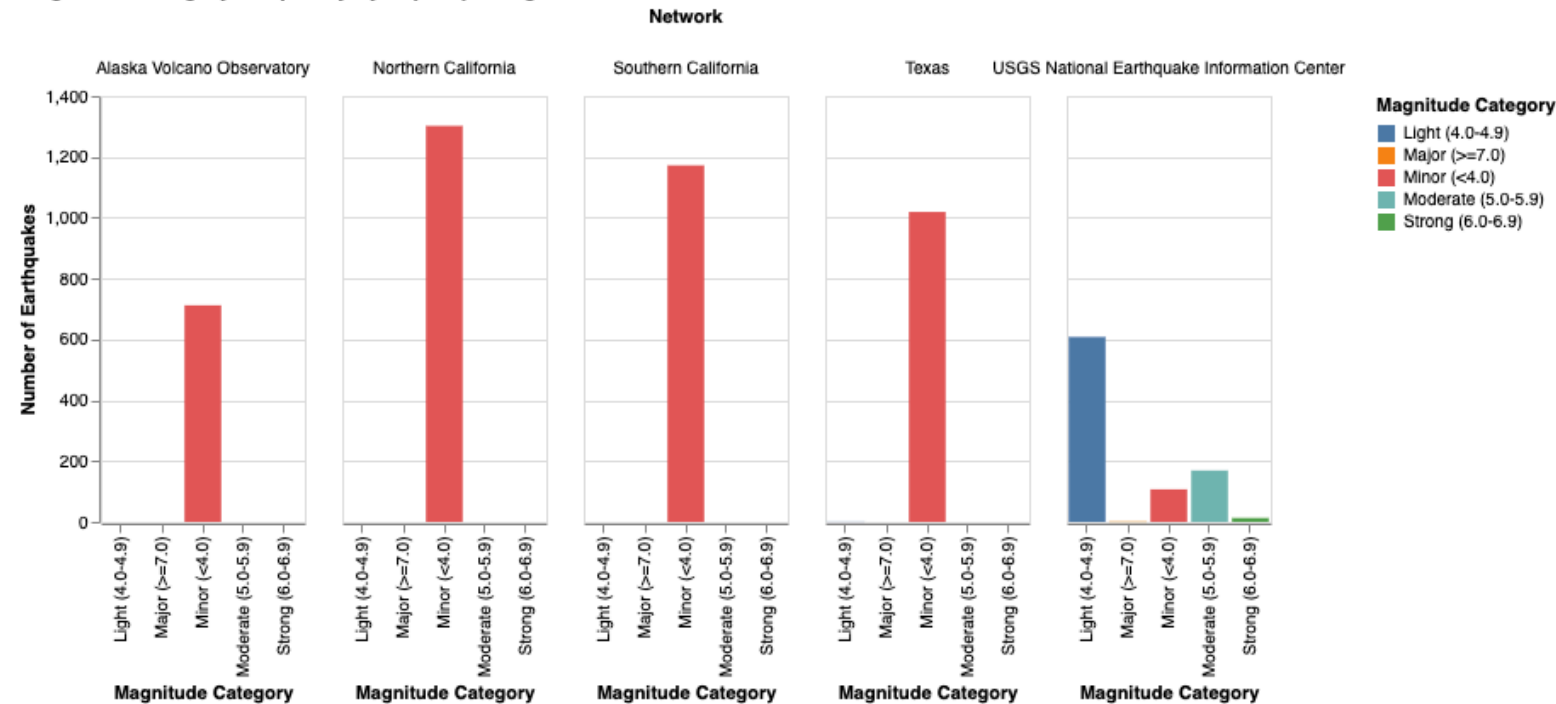
```
In [15]: # filter dataset to only top networks
top_earthquakes = earthquakes[earthquakes['net'].isin(top_nets)].copy()

network_labels = {
    'av': 'Alaska Volcano Observatory',
    'ci': 'Southern California',
    'nc': 'Northern California',
    'tx': 'Texas',
    'us': 'USGS National Earthquake Information Center'
}

top_earthquakes['net_full'] = top_earthquakes['net'].map(network_labels)

alt.Chart(top_earthquakes).mark_bar().encode(
    x=alt.X('mag_ordinal:N', title='Magnitude Category'),
    y=alt.Y('count()', title='Number of Earthquakes'),
    color=alt.Color('mag_ordinal:N', title='Magnitude Category'),
    column=alt.Column('net_full:N', title='Network')
).properties(width=120, height=250, title='Magnitude Category Frequency by Top Reporting Networks')
```

Out[15]: **Magnitude Category Frequency by Top Reporting Networks**



- Comparing magnitude categories across major networks shows how reporting differs regionally.
- For example, California networks (NC, CI) have the highest volumes of Minor events.
- In contrast, USGS reports fewer but more varying events, possibly due to detection thresholds or location.

E. Exploring Magnitude Calculations ( magType )

```
In [16]: unique = earthquakes['magType'].unique()
countunique = earthquakes['magType'].value_counts()

print("Unique magnitude types:")
print(unique)
print("\nCount of magnitude types:")
print(countunique)
```

```
Unique magnitude types:
['ml' 'md' 'mb' 'mwr' 'mww' 'mb_lg' 'mlv']

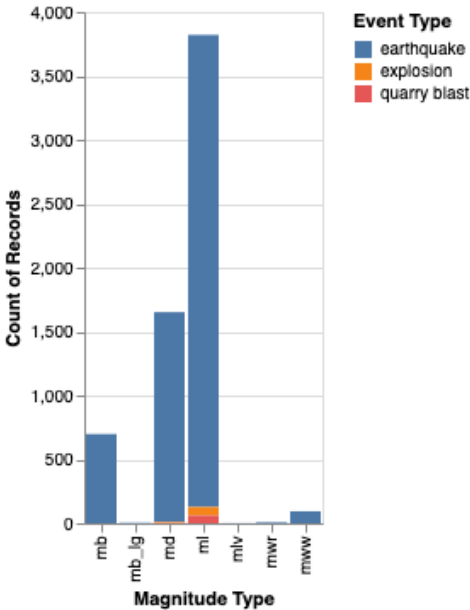
Count of magnitude types:
magType
ml      3821
md      1652
mb       698
mww       94
mwr        9
mb_lg      5
mlv        2
Name: count, dtype: int64
```

- `magType` indicates how magnitude was measured, and that the type used depends on earthquake size, distance, and instrument setup.
- Specifically, it specifies the method used to calculate an earthquake’s magnitude. The website details every single one [here](#), but only seven are included here: `ml` (local), `md` (duration), `mb` (short-period body wave), `mww` (moment w-phase), `mwr` (regional), `mb_lg` (short-period surface area), `mlv` (not mentioned, but likely a local variation of ML reported by a specific regional network),
- Different types are used depending on event size, distance, and instrument type. For instance, `ml` (Local Magnitude) measures nearby small earthquakes using the original Richter method, `md` is based on shaking duration for very small or clipped events, and `mb` or `mww/mwr` are used for larger or deeper global earthquakes.
- These distinctions help explain why average magnitudes and depths differ across magnitude types.

Distribution of Magnitude Type by Event Type

```
In [17]: alt.Chart(earthquakes).mark_bar().encode(
    alt.X('magType:N', title='Magnitude Type'),
    alt.Y('count():Q'),
    alt.Color('type:N', title='Event Type'),
    tooltip=['magType', 'type', 'count()']
).properties(title='Distribution of Magnitude Type by Event Type')
```

Out [17]: **Distribution of Magnitude Type by Event Type**

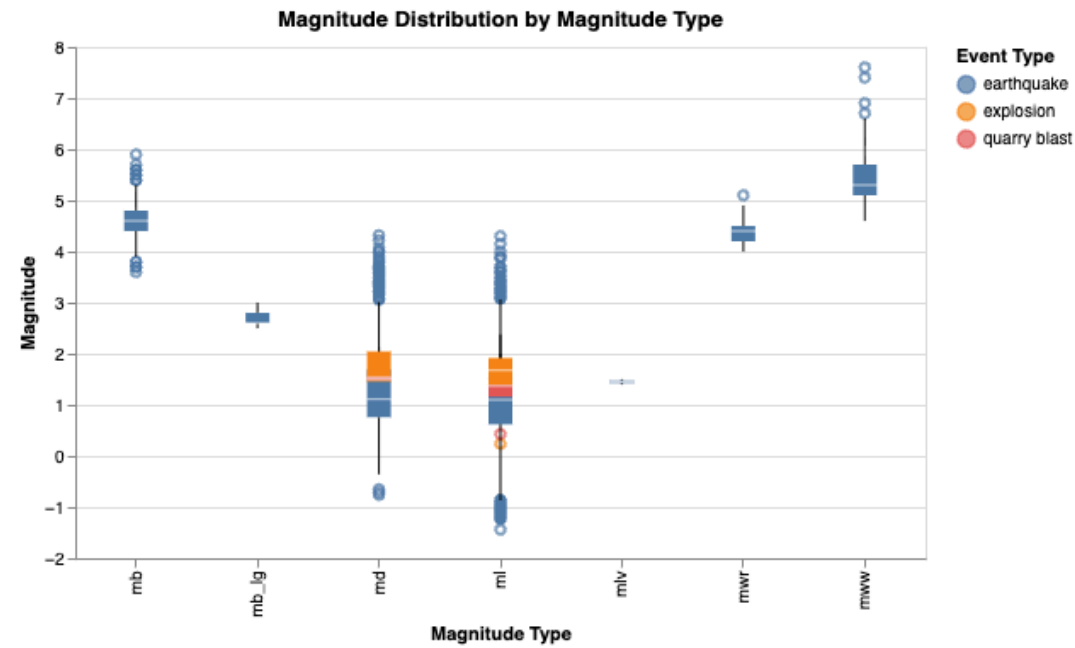


- This chart shows how frequently each `magType` occurs and for what event type.
- We can see most events were reported using local `ml` and duration `md` scales, with `ml` being the most common, which aligns with their definitions as local and duration, as they are based measures for small, nearby earthquakes, and as we saw earlier, more minor earthquakes (smaller magnitude) are more common in this dataset.
- Higher-magnitude methods like `mb`, `mb_lg`, `mwr`, and `mww` appear much less frequently, reflecting their use in larger events, which also corresponds to earlier findings of few higher-magnitude events.

```
In [18]: alt.Chart(earthquakes).mark_boxplot(extent=1.5).encode(  
    alt.X('magType:N', title='Magnitude Type'),  
    alt.Y('mag:Q', title='Magnitude'),  
    alt.Color('type:N', title='Event Type'),  
    tooltip=['mag', 'magType']  
) .properties(width=500, height=300, title='Magnitude Distribution by Magnitude Type')
```



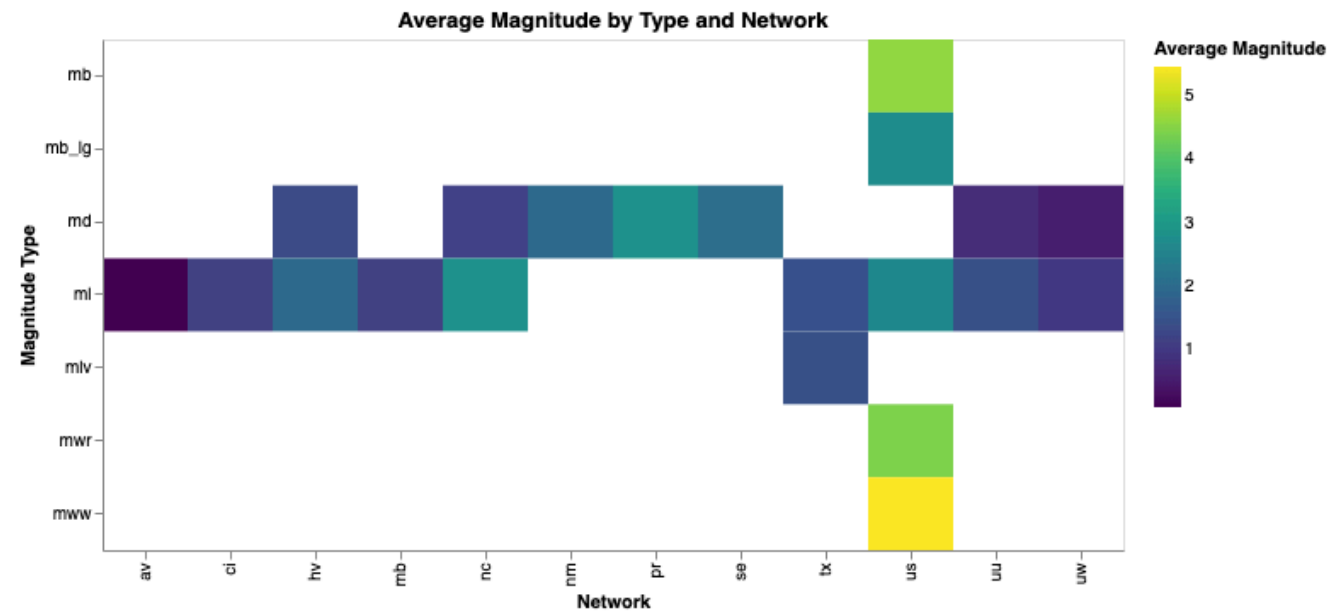
Out[18]:



```
In [19]: agg2 = earthquakes.groupby(['magType', 'net'], as_index=False)['mag'].mean()

alt.Chart(agg2).mark_rect().encode(
    alt.X('net:N', title='Network'),
    alt.Y('magType:N', title='Magnitude Type'),
    alt.Color('mag:Q', title='Average Magnitude', scale=alt.Scale(scheme='viridis')),
    tooltip=['mag', 'net', 'magType']
).properties(title='Average Magnitude by Type and Network', width=600, height=300)
```

Out [19] :



- This heatmap explores average magnitude varies across the different recording networks and magnitude measurement types.
- The U.S. network ( us ) records some of the highest average magnitudes, particularly for mww , mb , and mwr , whereas local networks such as ci and nc (Northern and Southern California) primarily record lower to moderate magnitudes, typically using local magnitude scales like ml or md .
- Magnitude types ml and md show a broader spread across multiple networks, suggesting these scales are commonly used for regional and smaller events.

F. Summary of EDA

- Most seismic events are shallow minor earthquakes (magnitude < 4, depth < 70 km).
- The correlation between magnitude and depth (r = 0.397) is weakly positive but suggests that deeper quakes can be moderately stronger, though variability is high.
- Regional networks influence observed patterns - local networks capture more events and more low-magnitude events, while USGS reports fewer, higher-magnitude events.
- Ordinal groupings show that all major earthquakes (≥ 7.0) are shallow.
- Magnitude types ( magType ) explain measurement differences.

In [ ] :