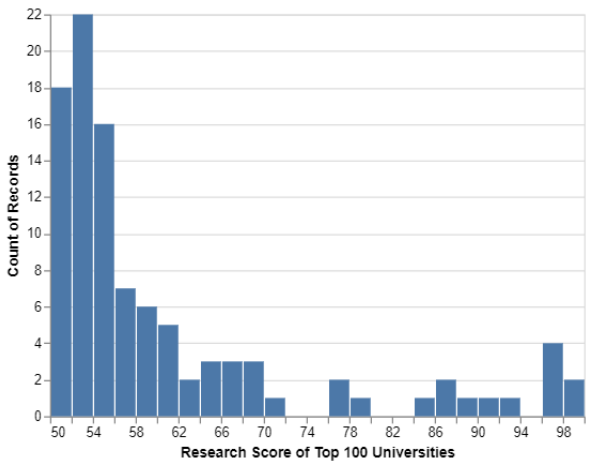
```
# Import the necessary libraries
import altair as alt
import pandas as pd

df = pd.read_excel("/content/WorldUniversity.xlsx")
df.head()
```

	world_rank	institution	country	national_rank	quality_of_education	alumni_employment	quality_of_faculty	publications	infl
0	1	Harvard University	USA	1	7	9	1	1	
1	2	Massachusetts Institute of Technology	USA	2	9	17	3	12	
2	3	Stanford University	USA	3	17	11	5	4	
3	4	University of Cambridge	United Kingdom	1	10	24	4	16	
4	5	California Institute of Technology	USA	4	2	29	7	37	

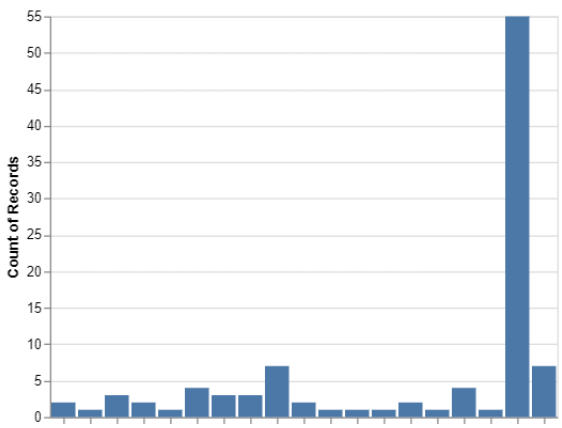
```
# to count year column and give information of data

df['year'].value_counts()
```

```
2014    1000
2015    1000
2012     100
2013     100
Name: year, dtype: int64
```

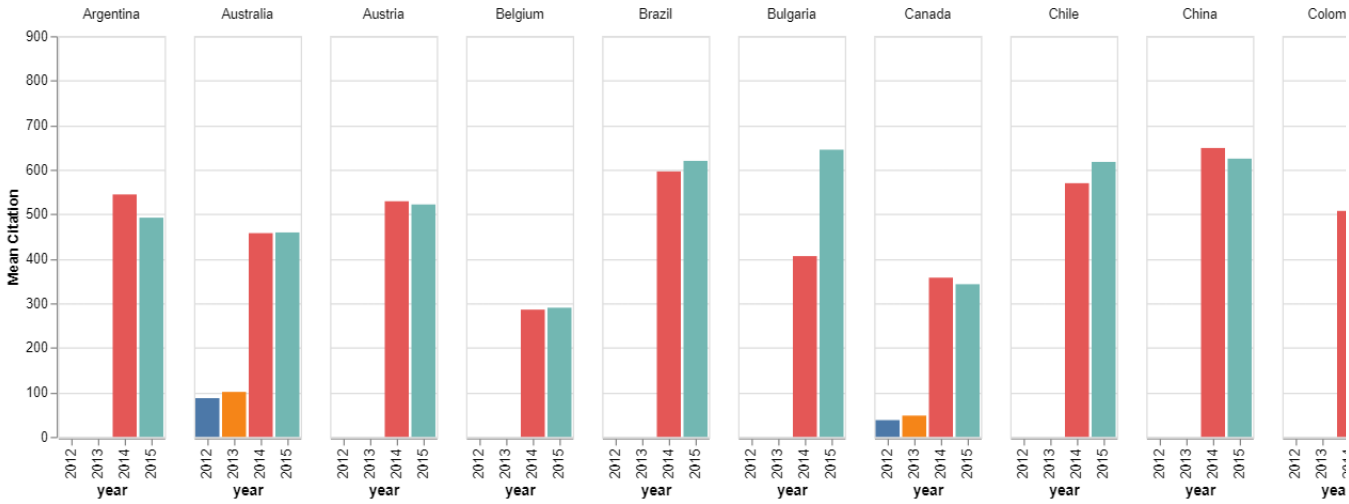
```
#a histogram of the research score of the top 100 universities.
df_latest = df[df['year']==df['year'].max()].reset_index(drop=True)

alt.Chart(df_latest.loc[:100]).mark_bar().encode(
    x = alt.X('score', bin = alt.BinParams(maxbins = 30),
              title="Research Score of Top 100 Universities"),

    y = 'count()'
)
```



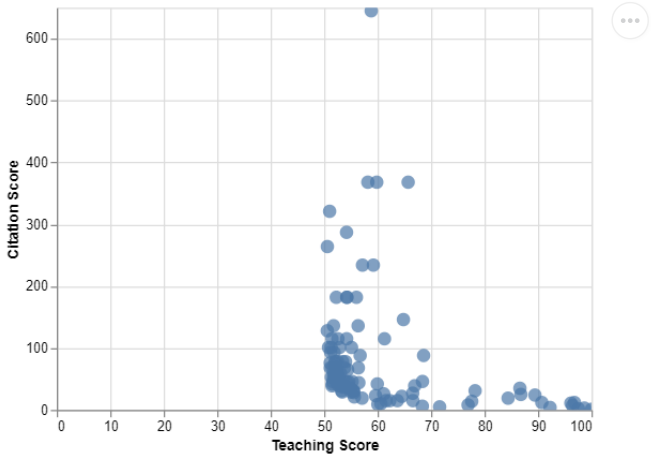

```
#barchart
alt.Chart(df_latest.loc[:100]).mark_bar().encode(
    x = 'country',
    y = 'count()'
)
```



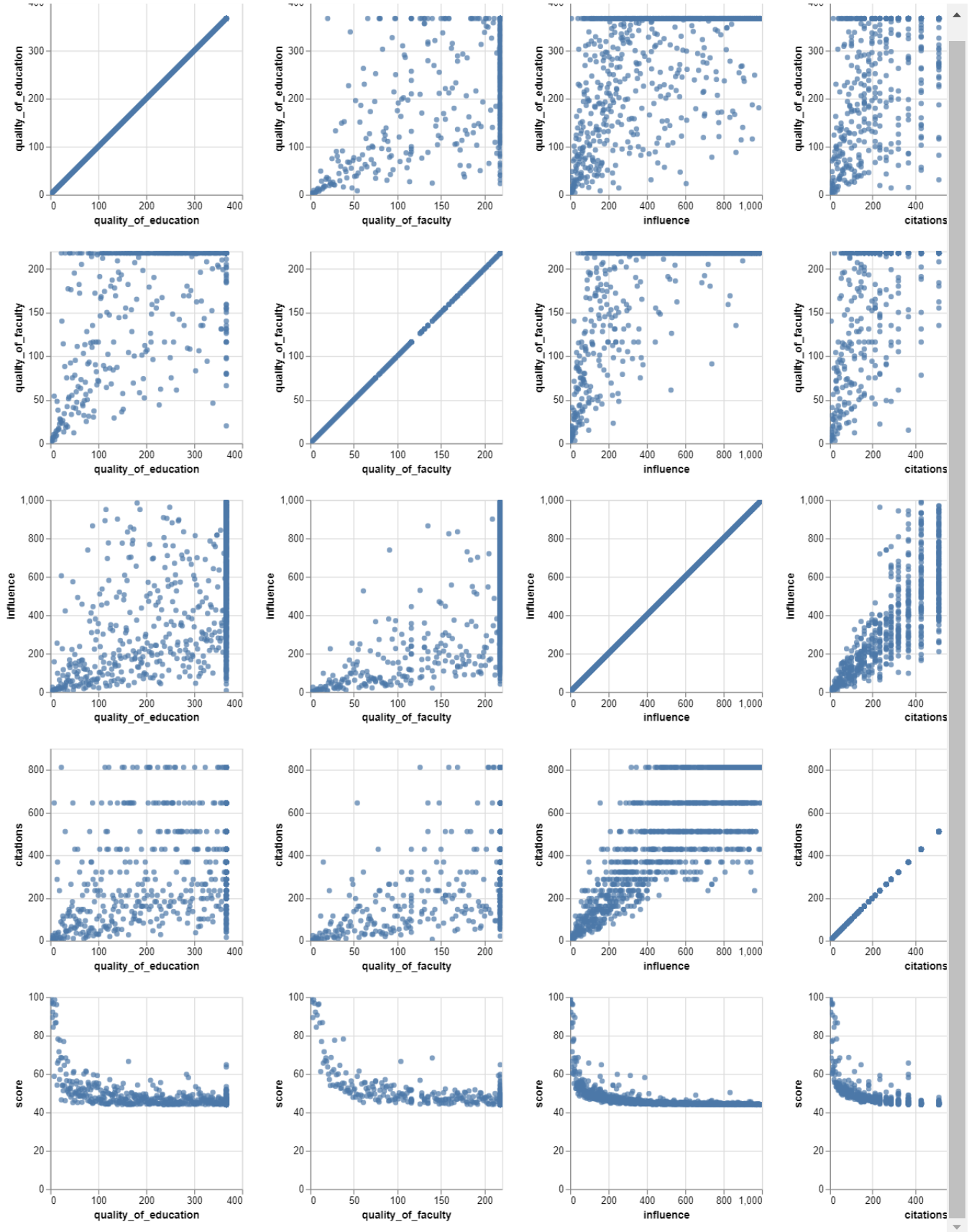

```
#group chart
bars = alt.Chart(df).mark_bar().encode(
        x='year:O',
        y=alt.Y('mean(citations):Q', title='Mean Citation'),
        color='year:N',
    )

alt.layer(bars).facet(
    column='country'
)
```



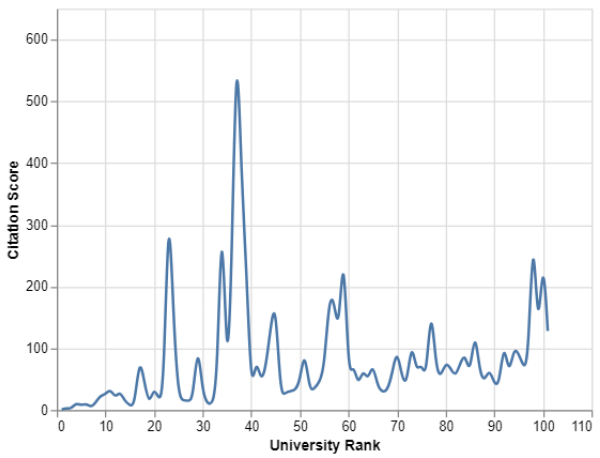

```
#scatter plot
alt.Chart(df_latest.loc[:100]).mark_circle(size=100).encode(
    x=alt.Y('score', title = "Teaching Score"),
    y=alt.Y('citations', title = "Citation Score"),
    tooltip=['score', 'citations']
).interactive()
```



```
values_n = ['quality_of_education','quality_of_faculty','influence','citations','score']

alt.Chart(df_latest).mark_circle().encode(

    alt.X(alt.repeat("column"), type='quantitative'),
    alt.Y(alt.repeat("row"), type='quantitative')
).properties(
    width=200,
    height=200
).repeat(
    row=(values_n),
    column=(values_n)
).interactive()
```



```
#line chart
alt.Chart(df_latest.loc[:100,:]).mark_line(interpolate='basis').encode(
    x = alt.X('world_rank:Q', title = "University Rank"),
    y = alt.Y('citations:Q', title = "Citation Score")
)
```




```
#hybrid chart
heatmap = alt.Chart(df_latest.loc[:100,:]).mark_rect().encode(
    alt.X('citations:Q', bin=alt.BinParams(maxbins = 20), title = "Teaching Score"),
    alt.Y('score:Q', bin=alt.BinParams(maxbins = 20), title = "Research Score"),
    alt.Color('count():Q', scale=alt.Scale(scheme='greenblue'))
)
points = alt.Chart(df_latest.loc[:100,:]).mark_circle(color='black',size=10).encode(
    x='citations:Q',
    y='score:Q',
)

heatmap + points
```