Billionaires Statistics Data Analysis and Visualization

Import Libraries

In [1]:
```
import plotly.express as px
import plotly.graph_objects as go
import pandas as pd
```

In [2]:
```
df = pd.read_csv("/input/billionaires-statistics/Billionaires Statistics Dataset.csv")
df
```

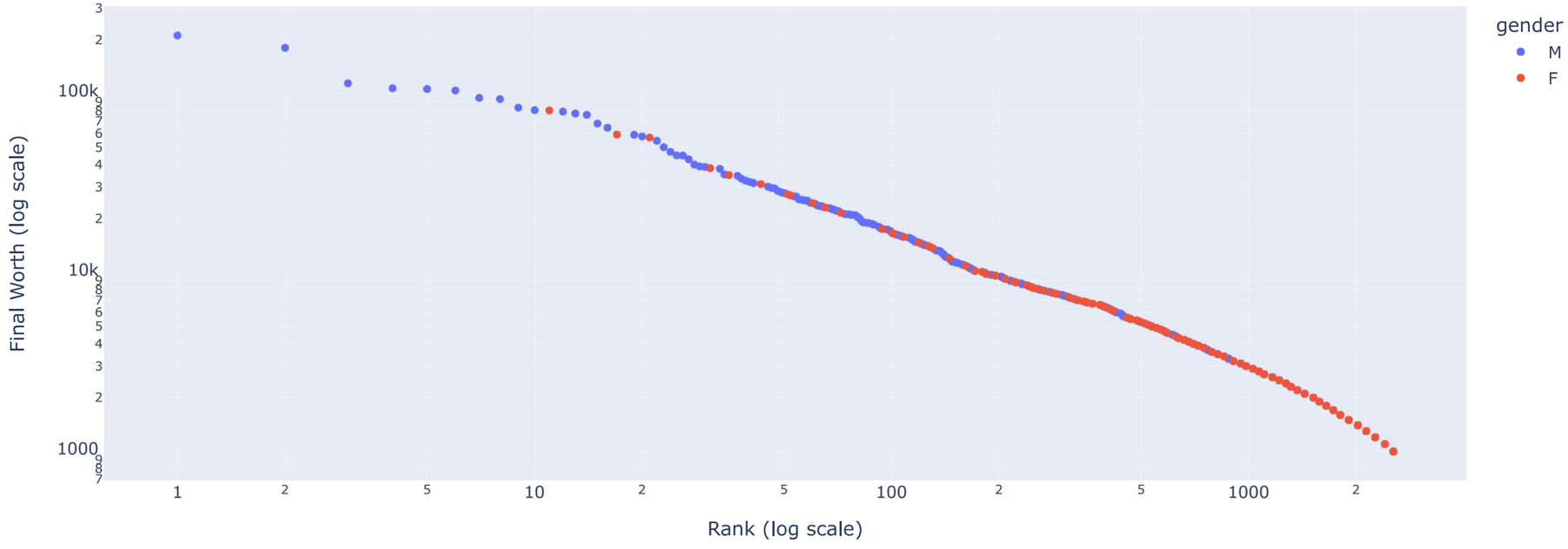
Out[2]:

	rank	finalWorth	category	personName	age	country	city	source	industries	countryOfCitizenship	...	cpi_change_country	gdp_country	gross_tertiary_education_enrollment	gross_primary_education_
0	1	211000	Fashion & Retail	Bernard Arnault & family	74.0	France	Paris	LVMH	Fashion & Retail	France	...	1.1	$2,715,518,274,227	65.6	
1	2	180000	Automotive	Elon Musk	51.0	United States	Austin	Tesla, SpaceX	Automotive	United States	...	7.5	$21,427,700,000,000	88.2	
2	3	114000	Technology	Jeff Bezos	59.0	United States	Medina	Amazon	Technology	United States	...	7.5	$21,427,700,000,000	88.2	
3	4	107000	Technology	Larry Ellison	78.0	United States	Lanai	Oracle	Technology	United States	...	7.5	$21,427,700,000,000	88.2	
4	5	106000	Finance & Investments	Warren Buffett	92.0	United States	Omaha	Berkshire Hathaway	Finance & Investments	United States	...	7.5	$21,427,700,000,000	88.2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2635	2540	1000	Healthcare	Yu Rong	51.0	China	Shanghai	Health clinics	Healthcare	China	...	2.9	$19,910,000,000,000	50.6	
2636	2540	1000	Food & Beverage	Richard Yuengling, Jr.	80.0	United States	Pottsville	Beer	Food & Beverage	United States	...	7.5	$21,427,700,000,000	88.2	
2637	2540	1000	Manufacturing	Zhang Gongyun	60.0	China	Gaomi	Tyre manufacturing machinery	Manufacturing	China	...	2.9	$19,910,000,000,000	50.6	
2638	2540	1000	Real Estate	Zhang Guiping & family	71.0	China	Nanjing	Real estate	Real Estate	China	...	2.9	$19,910,000,000,000	50.6	
2639	2540	1000	Diversified	Inigo Zobel	66.0	Philippines	Makati	Diversified	Diversified	Philippines	...	2.5	$376,795,508,680	35.5	

2640 rows × 35 columns

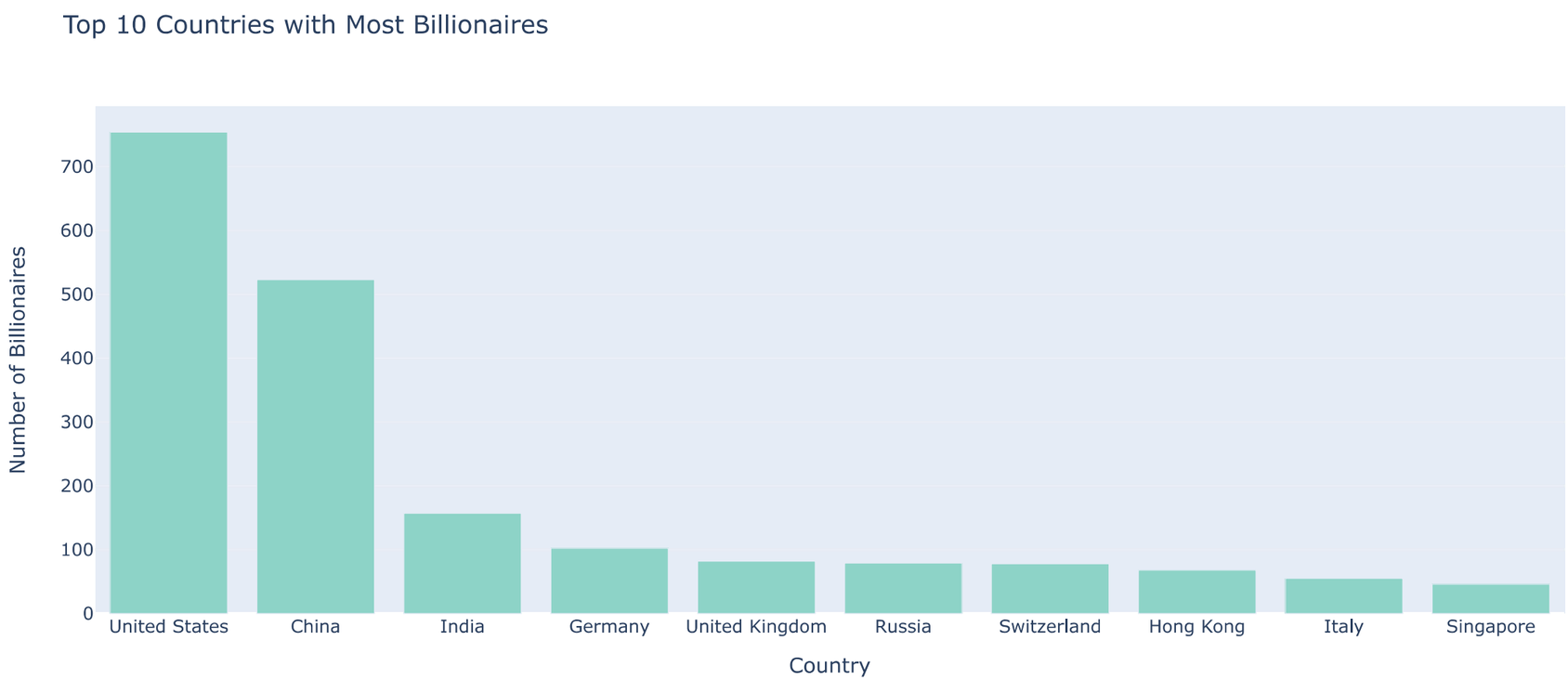
Billionaires' Rank vs. Final Worth by Gender

In [3]:
```
fig1 = px.scatter(df, x="rank", y="finalWorth", color="gender", title="Billionaires' Rank vs. Final Worth by Gender")
fig1.update_xaxes(type='log', title="Rank (log scale)")
fig1.update_yaxes(type='log', title="Final Worth (log scale)")
fig1.show()
```



Bar Chart: Top 10 Countries with Most Billionaires

In [4]:
```
top_countries = df['country'].value_counts().head(10)
fig2 = px.bar(top_countries, x=top_countries.index, y=top_countries.values,
              title="Top 10 Countries with Most Billionaires",
              color_discrete_sequence=px.colors.qualitative.Set3)
fig2.update_xaxes(title="Country")
fig2.update_yaxes(title="Number of Billionaires")
fig2.show()
```



Pie chart: Gender Distribution

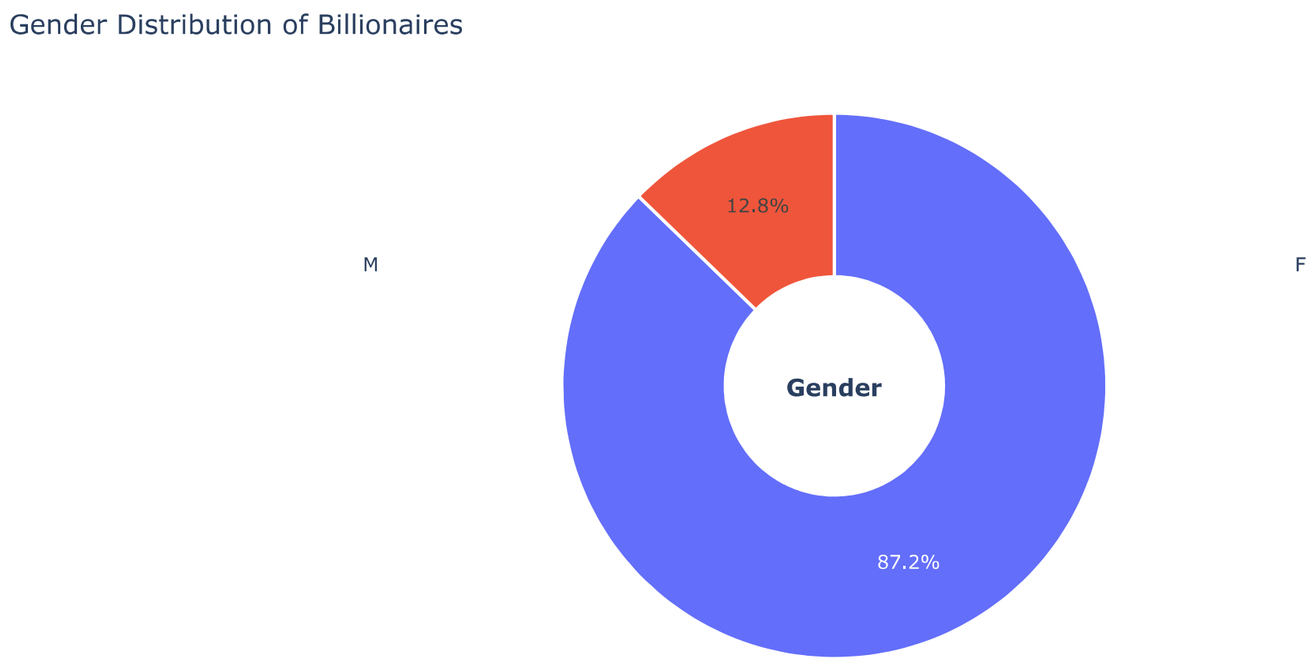
In [5]:
```
gender_counts = df['gender'].value_counts()

fig3 = px.pie(gender_counts, labels=gender_counts.index, values=gender_counts.values,
              title="Gender Distribution of Billionaires",
              color_discrete_sequence=px.colors.qualitative.Plotly)
fig3.update_traces(marker=dict(line=dict(color='white', width=2)))
fig3.update_layout(showlegend=False)
fig3.update_traces(hole=0.4)

gender_labels = gender_counts.index
fig3.add_annotation(
    text="<b>Gender</b>",
    x=0.5,
    y=0.5,
    showarrow=False,
    font=dict(size=15),
)
fig3.add_annotation(
    text=gender_labels[0],
    x=0.2,
    y=0.75,
    showarrow=False,
    font=dict(size=12),
```

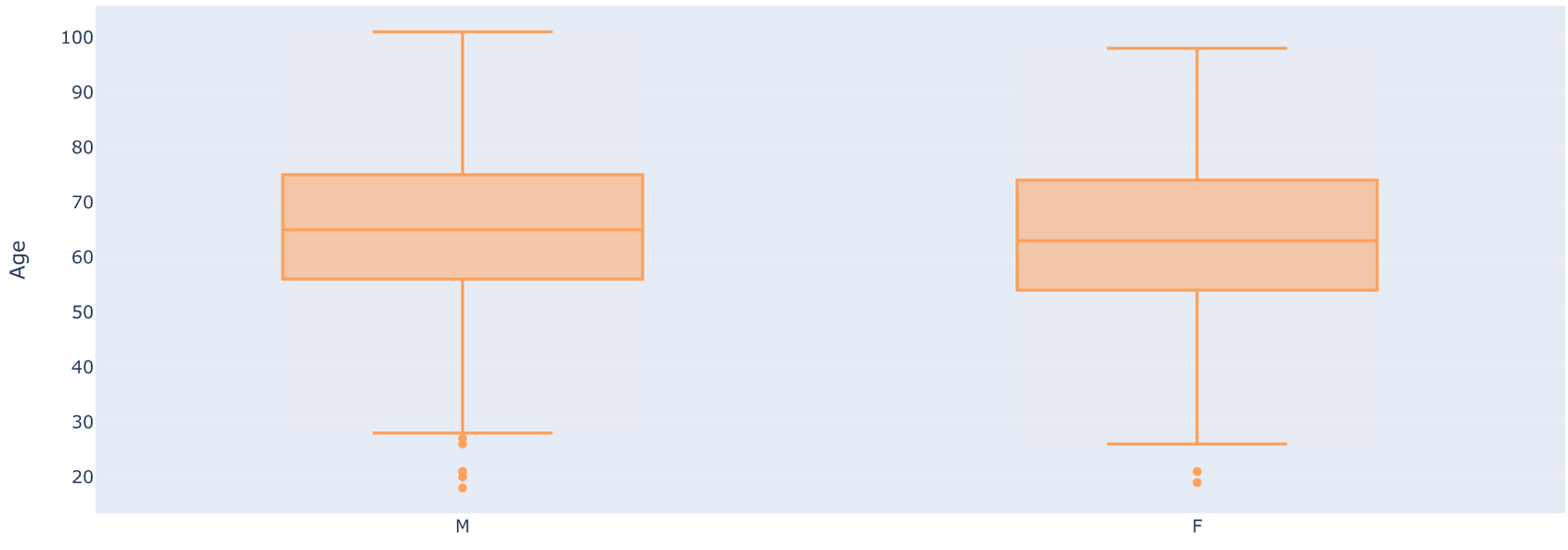
```
)
fig3.add_annotation(
    text=gender_labels[1],
    x=0.8,
    y=0.75,
    showarrow=False,
    font=dict(size=12),
)

fig3.show()
```



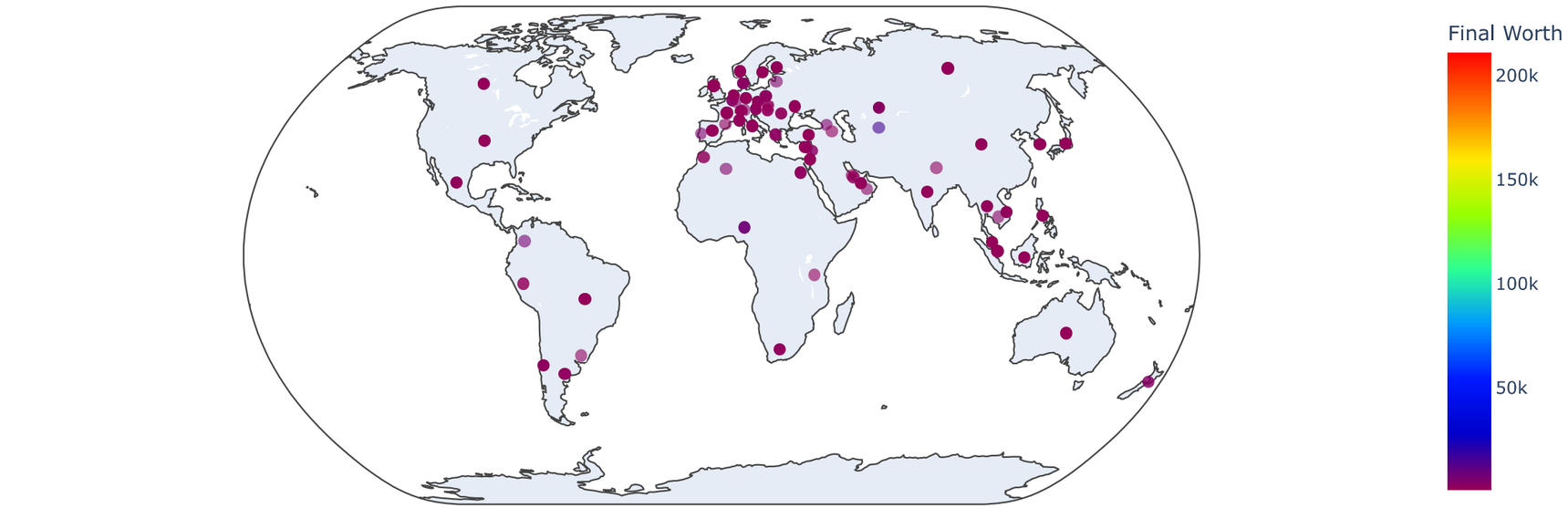
Box plot: Age Distribution by Gender

In [6]:
```
fig4 = px.box(df, x="gender", y="age", title="Age Distribution of Billionaires by Gender",
              color_discrete_sequence=['#FFA15A', '#00B2E2'])
fig4.update_xaxes(title="Gender")
fig4.update_yaxes(title="Age")
fig4.show()
```



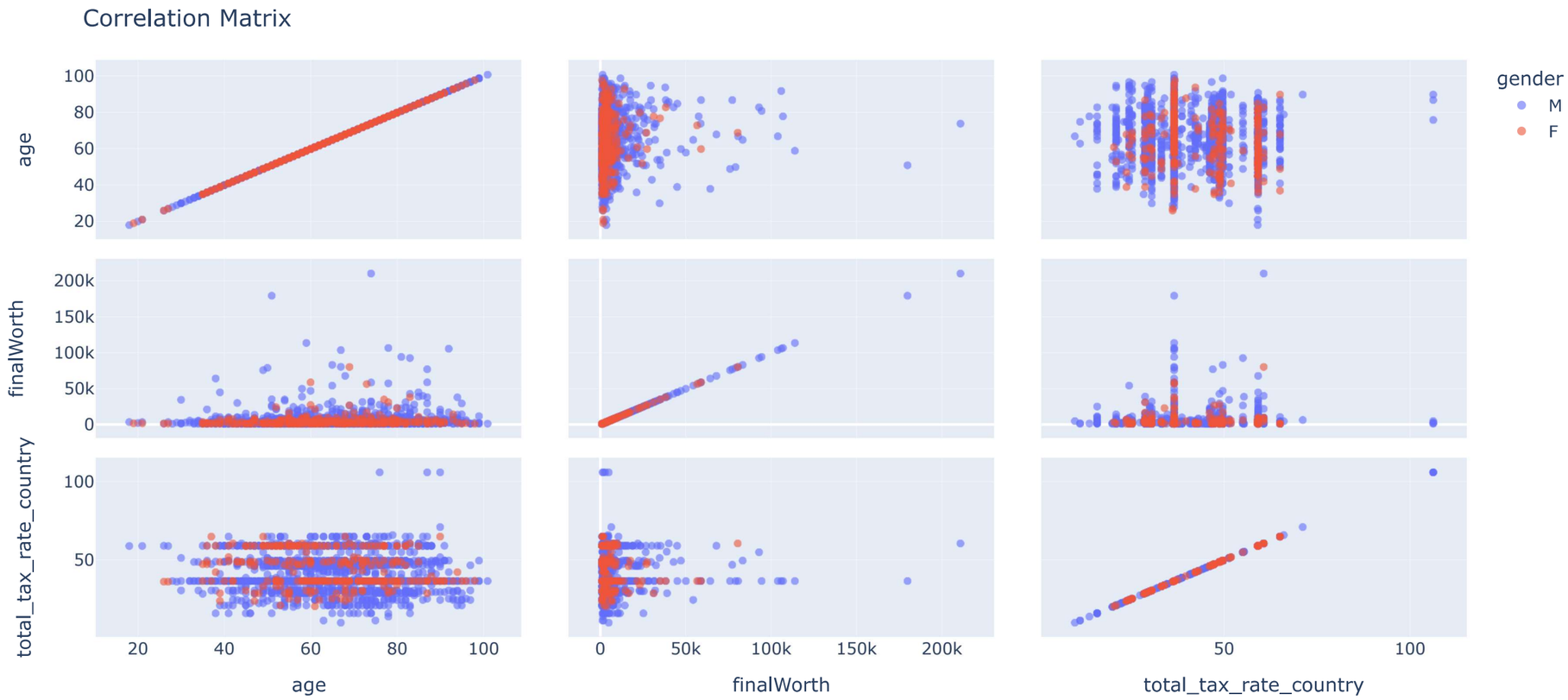
Geospatial plot: Billionaires' Distribution on World Map

In [7]:
```
fig5 = go.Figure(data=go.Scattergeo(
    lon=df["longitude_country"],
    lat=df["latitude_country"],
    text=df["personName"],
    mode="markers",
    marker=dict(
        size=8,
        opacity=0.6,
        color=df["finalWorth"],
        colorscale="Rainbow",
        colorbar=dict(title="Final Worth")
    )
))
fig5.update_geos(projection_type="natural earth")
fig5.update_layout(title="Billionaires' Distribution on World Map")
fig5.show()
```



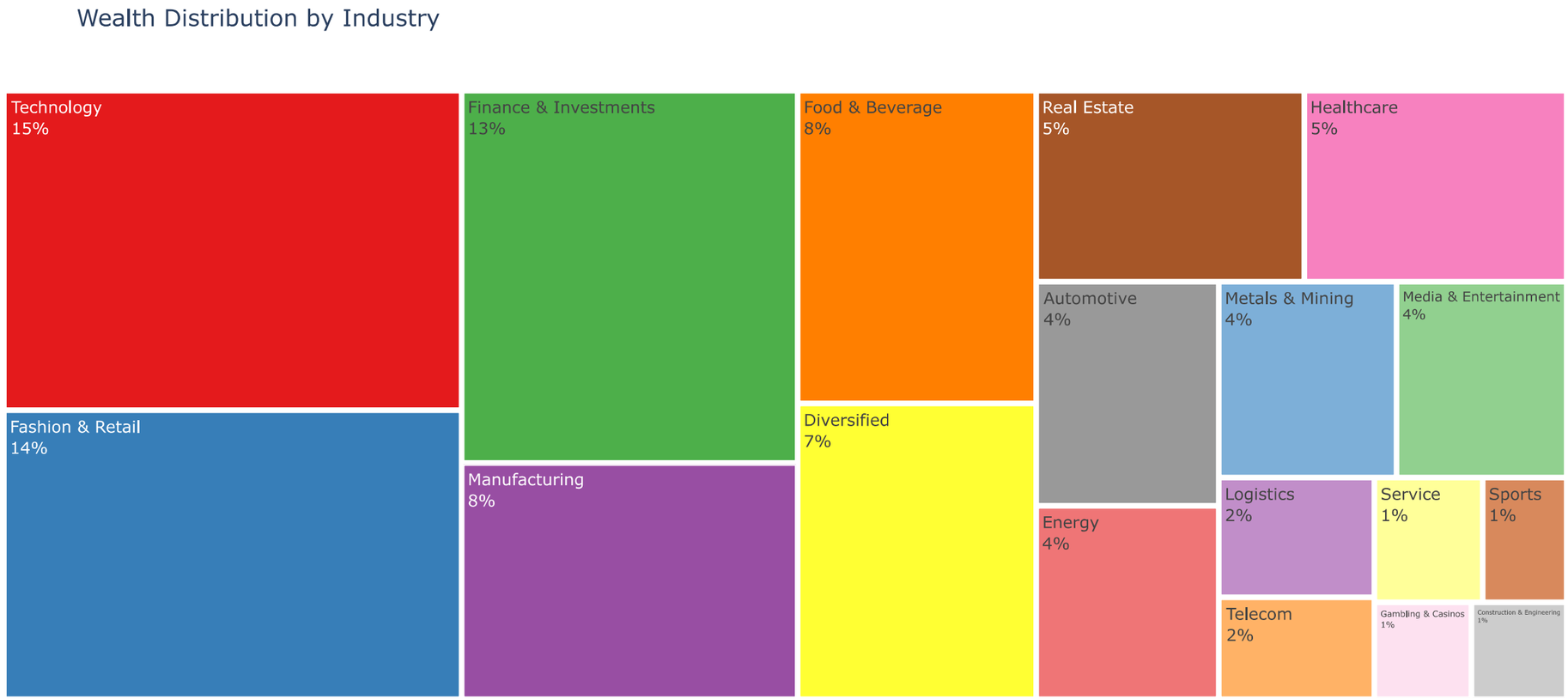
Scatter Matrix: Correlation between Age, Final Worth, and Tax Rate

In [8]:
```
fig6 = px.scatter_matrix(df, dimensions=["age", "finalWorth", "total_tax_rate_country"],
                         color="gender", title='Correlation Matrix')
fig6.update_traces(marker=dict(size=6, opacity=0.6))
fig6.update_layout(margin=dict(t=50, l=50, r=50, b=50))
fig6.show()
```



Treemap: Wealth Distribution by Industry

In [9]:
```
fig7 = px.treemap(df, path=['industries'], values='finalWorth',
                  title='Wealth Distribution by Industry',
                  color_discrete_sequence=px.colors.qualitative.Set1)
fig7.update_traces(textinfo="label+percent entry")
fig7.update_layout(margin=dict(t=50, l=0, r=0, b=0))
fig7.show()
```



Comparison of Wealth by Age and Gender (Radar Chart)

In [10]:
```
fig8 = px.line_polar(df, r='finalWorth', theta='age', color='gender',
                     title='Comparison of Wealth by Age and Gender (Radar Chart)',
                     line_close=True, color_discrete_sequence=px.colors.qualitative.Plotly)
fig8.update_traces(fill='toself')
fig8.show()
```

Comparison of Wealth by Age and Gender (Radar Chart)