Principle Component Analysis

In [4]:
```
import plotly.express as px
from sklearn.decomposition import PCA

df = px.data.iris()
features = ["sepal_width", "sepal_length", "petal_width", "petal_length"]

pca = PCA()
components = pca.fit_transform(df[features])
labels = {
    str(i): f"PC {i+1} ({var:.1f}%)"
    for i, var in enumerate(pca.explained_variance_ratio_ * 100)
}

fig = px.scatter_matrix(
    components,
    labels=labels,
    dimensions=range(4),
    color=df["species"]
)
fig.update_traces(diagonal_visible=False)
fig.show()
```

In [5]:
```
import numpy as np
exp_var_cumul = np.cumsum(pca.explained_variance_ratio_)

px.area(
    x = range(1, exp_var_cumul.shape[0] + 1),
    y = exp_var_cumul,

    labels = {"x": "# Components", "y": "Explained Variance"}
)
```

In [ ]:
```

```