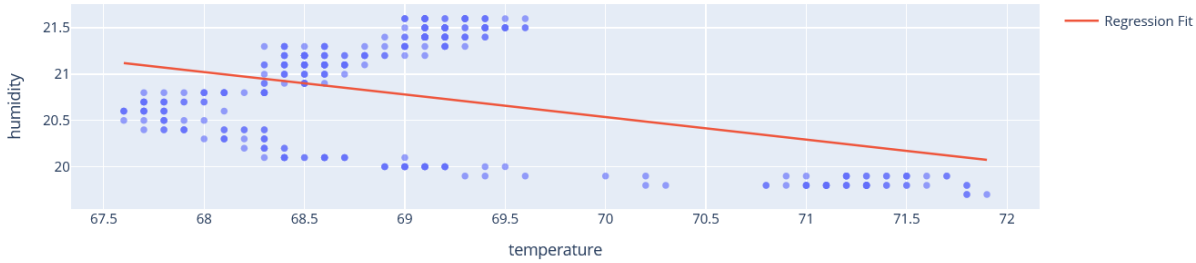
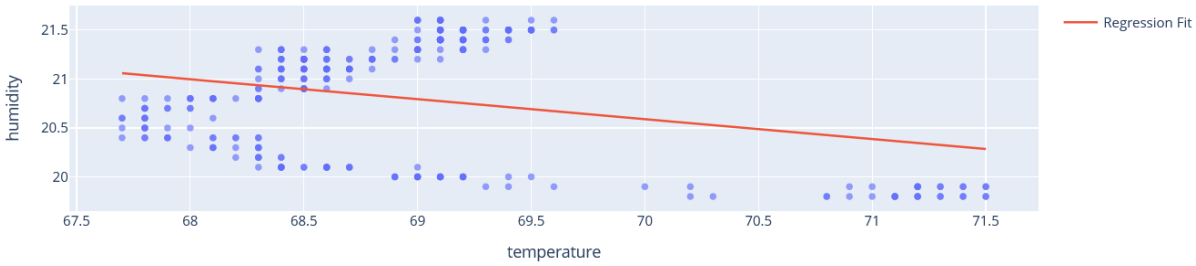
```
import pandas as pd
import numpy as np
import plotly.express as px
import plotly.graph_objects as go
from sklearn.linear_model import LinearRegression

df = pd.read_csv("7.1P_data.csv")

X = df.temperature.values.reshape(-1, 1)

model = LinearRegression()
model.fit(X, df.humidity)

x_range = np.linspace(X.min(), X.max(), 100)
y_range = model.predict(x_range.reshape(-1, 1))

fig = px.scatter(df, x='temperature', y='humidity', opacity=0.65, title="First Comparision")
fig.add_traces(go.Scatter(x=x_range, y=y_range, name='Regression Fit'))
fig.show()

#remove 10 max and min values each for next cell
df = df.sort_values('temperature')
df = df.drop(df.tail(10).index)
df = df.drop(df.head(10).index)
df = df.sort_values('humidity')
df = df.drop(df.tail(10).index)
df = df.drop(df.head(10).index)
```

First Comparision




```
#repeat to train model to new df
X = df.temperature.values.reshape(-1, 1)

model.fit(X, df.humidity)

x_range = np.linspace(X.min(), X.max(), 100)
y_range = model.predict(x_range.reshape(-1, 1))

fig = px.scatter(df, x='temperature', y='humidity', opacity=0.65, title="Training with Outliers Removed")
fig.add_traces(go.Scatter(x=x_range, y=y_range, name='Regression Fit'))
fig.show()
```

Training with Outliers Removed



```
#repeat to remove more values
df = df.sort_values('temperature')
df = df.drop(df.tail(10).index)
df = df.sort_values('humidity')
df = df.drop(df.tail(10).index)

X = df.temperature.values.reshape(-1, 1)

model.fit(X, df.humidity)

x_range = np.linspace(X.min(), X.max(), 100)
y_range = model.predict(x_range.reshape(-1, 1))

fig = px.scatter(df, x='temperature', y='humidity', opacity=0.65, title="Removing More Outliers")
fig.add_traces(go.Scatter(x=x_range, y=y_range, name='Regression Fit'))
fig.show()
```