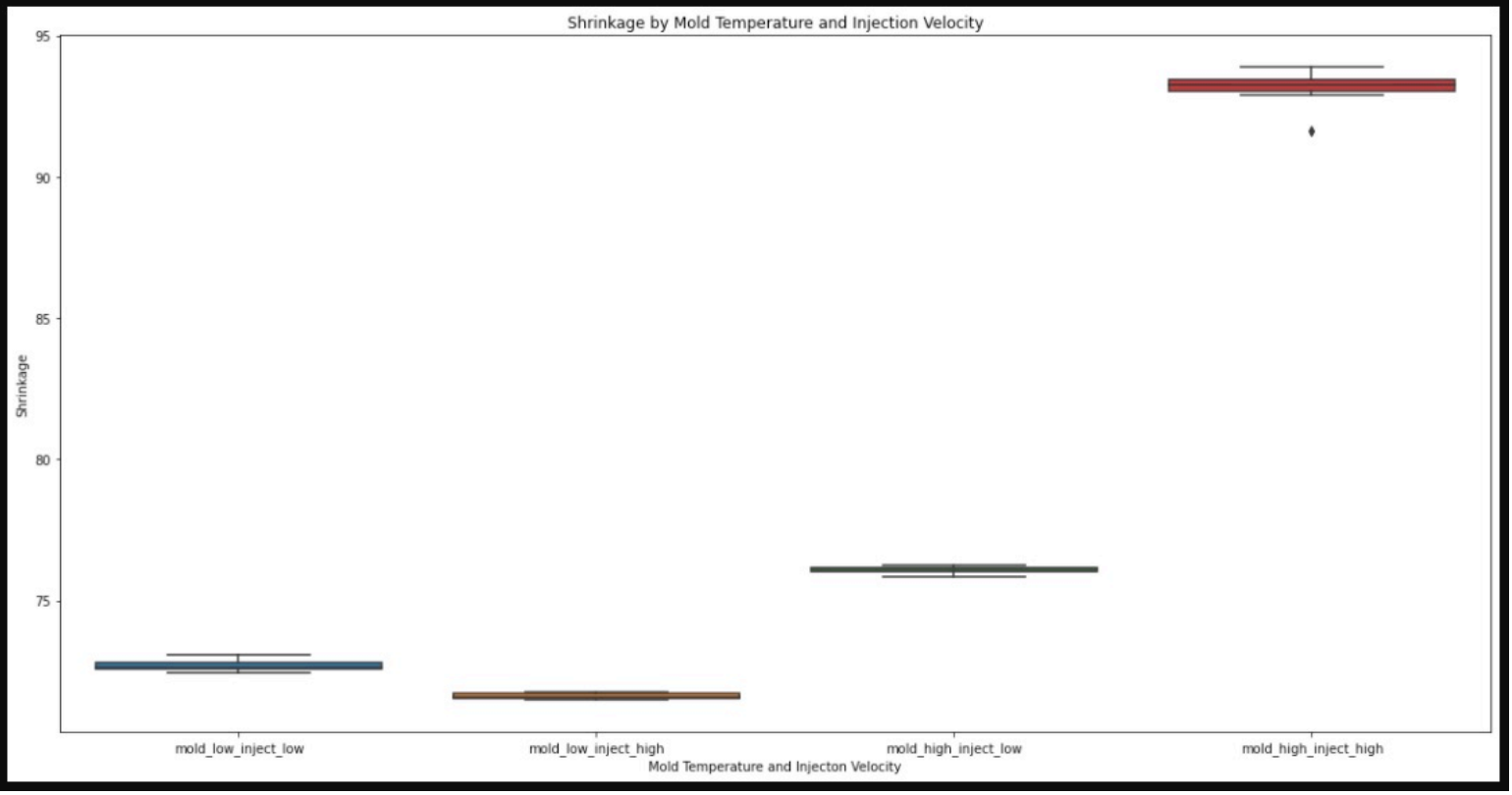
Problem 1.32

Use the results of Exercises 1.28 and 1.31 to create a plot that illustrates the interaction evident from the data. Use the plot in Figure 1.3 in Example 1.3 as a guide. Could the type of information found in Exercises 1.28 and 1.31 have been found in an observational study in which there was no control on injection velocity and mold temperature by the analyst? Explain why or why not.

```
plt.figure(figsize=(20, 10))

prob_1_33_df = pd.DataFrame({'mold_low_inject_low':prob_1_28_low,
                             'mold_low_inject_high':prob_1_28_high,
                             'mold_high_inject_low':prob_1_31_low,
                             'mold_high_inject_high':prob_1_31_high})
ax = sns.boxplot(data=prob_1_33_df)
ax.set_title("Shrinkage by Mold Temperature and Injection Velocity")
ax.set_ylabel("Shrinkage")
ax.set_xlabel("Mold Temperature and Injecton Velocity")
```

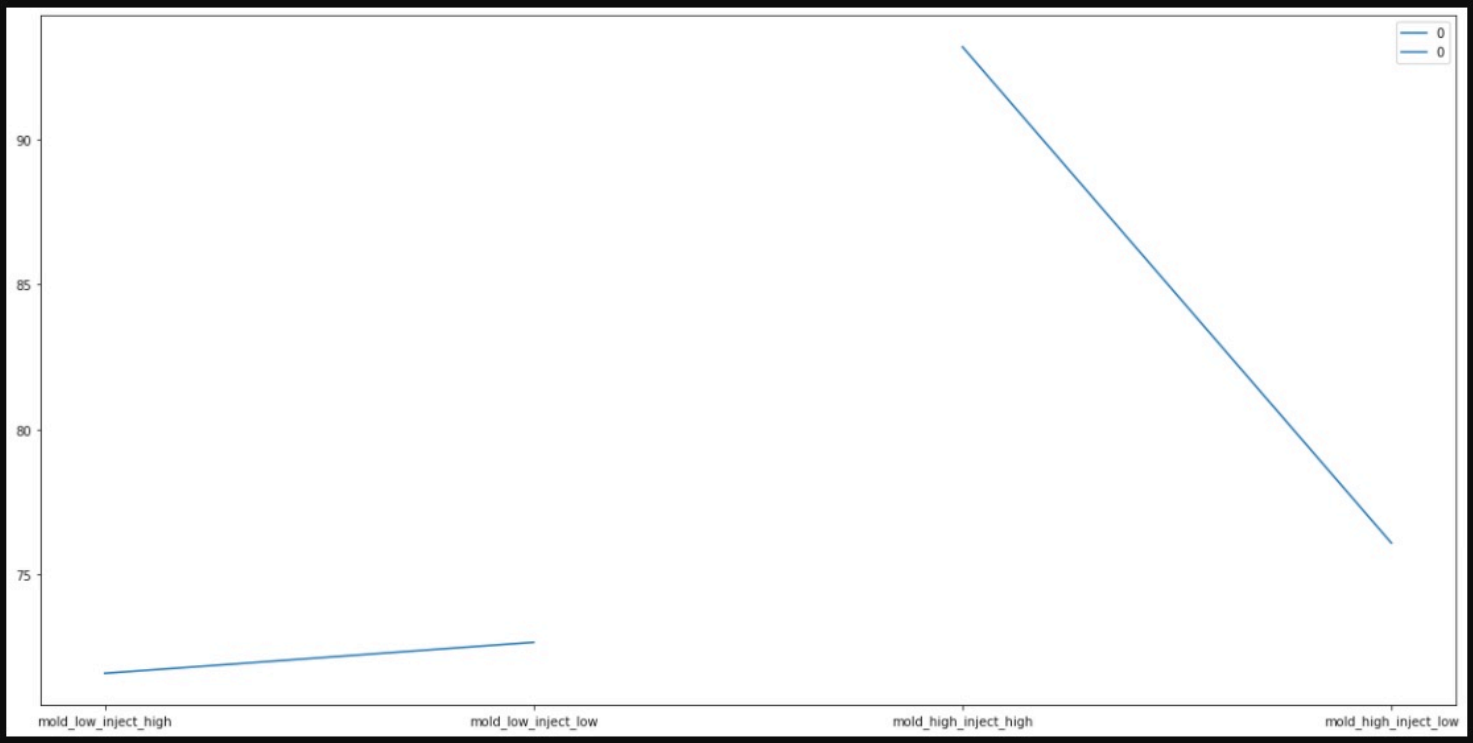
```
Text(0.5, 0, 'Mold Temperature and Injecton Velocity')
```



[191]:
```
prob_1_33_df = pd.DataFrame([prob_1_28_low.mean(),
                             prob_1_28_high.mean()],
                            index=['mold_low_inject_low', 'mold_low_inject_high'])
prob_1_33_df_1 = pd.DataFrame([prob_1_31_low.mean(),
                             prob_1_31_high.mean()],
                            index=['mold_high_inject_low', 'mold_high_inject_high'])

plt.figure(figsize=(20, 10))
sns.lineplot(data=prob_1_33_df)
sns.lineplot(data=prob_1_33_df_1)
```

[191]: <AxesSubplot:>



[115]:
```
train_low_low_df = pd.DataFrame(prob_1_28_low, columns=['target'])
train_low_low_df['inject_low'] = 1
train_low_low_df['mold_low'] = 1
train_low_low_df['inject_high'] = 0
train_low_low_df['mold_high'] = 0

train_high_low_df = pd.DataFrame(prob_1_28_high, columns=['target'])
train_high_low_df['inject_low'] = 0
train_high_low_df['mold_low'] = 1
train_high_low_df['inject_high'] = 1
train_high_low_df['mold_high'] = 0

train_low_high_df = pd.DataFrame(prob_1_31_low, columns=['target'])
train_low_high_df['inject_low'] = 1
train_low_high_df['mold_low'] = 0
train_low_high_df['inject_high'] = 0
train_low_high_df['mold_high'] = 1

train_high_high_df = pd.DataFrame(prob_1_31_high, columns=['target'])
train_high_high_df['inject_low'] = 0
train_high_high_df['mold_low'] = 0
train_high_high_df['inject_high'] = 1
train_high_high_df['mold_high'] = 1


train_df = train_low_low_df\
    .append(train_high_low_df)\
    .append(train_low_high_df)\
    .append(train_high_high_df)
train_df.head()
```

[115]:

	target	inject_low	mold_low	inject_high	mold_high
0	72.68	1	1	0	0
1	72.62	1	1	0	0
2	72.58	1	1	0	0
3	72.48	1	1	0	0
4	73.07	1	1	0	0

166]:
```
from sklearn.linear_model import LinearRegression

independent_cols = ['inject_low','mold_low','inject_high','mold_high']
print(f'independent_cols: {independent_cols}\n')

regr = LinearRegression(fit_intercept = True, copy_X = True, n_jobs = 2)

X = train_df[independent_cols]
y = train_df['target']

regr.fit(X,y)
print(f'coeffs: {np.column_stack((independent_cols, regr.coef_))}\n')
regr.score(X,y)
```

```
independent_cols: ['inject_low', 'mold_low', 'inject_high', 'mold_high']

coeffs: [['inject_low' '-4.0024999999999995']
 ['mold_low' '-6.2444999999999995']
 ['inject_high' '4.0024999999999995']
 ['mold_high' '6.244499999999999']]
```

166]:
```
0.7267559993427677
```

add interaction term

171]:
```
independent_inter_cols_1 = independent_cols.copy()
independent_inter_cols_1.append('mold_low_inject_low')
print(f'independent_inter_cols_1: {independent_inter_cols_1}\n')

train_interaction_df = train_df.copy()
train_interaction_df['mold_low_inject_low'] = train_df['mold_low'] * train_df['inject_low']
X = train_interaction_df[independent_inter_cols_1]

regr = LinearRegression(fit_intercept = True, copy_X = True, n_jobs = 2)

regr.fit(X,y)
print(f'coeffs: {np.column_stack((independent_inter_cols_1, regr.coef_))}\n')
regr.score(X,y)
```

```
independent_inter_cols_1: ['inject_low', 'mold_low', 'inject_high', 'mold_high', 'mold_low_inject_low']

coeffs: [['inject_low' '-8.539000000000001']
 ['mold_low' '-10.780999999999997']
 ['inject_high' '8.538999999999994']
 ['mold_high' '10.780999999999995']
 ['mold_low_inject_low' '18.146000000000004']]
```

171]:
```
0.9986245163944938
```

Answer

The plots above show that there is an interaction between mold temperature and injection velocity. I dug deeper into the interaction using ordinary least squares regression. With no interaction term I get an R^2 of 0.73. After adding an interaction term R^2 increases to 0.998.

An observational study wouldn't provide the same clarity in the results. Shrinkage may be affected by more variables than mold temperature and injection velocity which would confound the results. The best way to identify specific relationships is to control variables of interest while holding everything else constant.