Impact of College Type and Region on Career Success

Setup

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
sns.set_context('talk')
import warnings
warnings.filterwarnings('ignore')

from scipy.stats import ttest_ind
```

College Type Data

In [2]:
```
college_type_df = pd.read_csv('data/salaries-by-college-type.csv')
```

In [3]:
```
college_type_df.head()
```

Out[3]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00	$76,800.00	$99,200.00	$168,000.00	$220,000.00
1	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00	NaN	$104,000.00	$161,000.00	NaN
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00	NaN	$96,000.00	$180,000.00	NaN
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00	$66,800.00	$94,300.00	$143,000.00	$190,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00	NaN	$80,200.00	$142,000.00	NaN

In [4]:
```
college_type_df.shape
```

Out[4]: (269, 8)

Region Data

In [5]:
```
region_df = pd.read_csv('data/salaries-by-region.csv')
```

In [6]:
```
region_df.head()
```

Out[6]:

	School Name	Region	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	Stanford University	California	$70,400.00	$129,000.00	$68,400.00	$93,100.00	$184,000.00	$257,000.00
1	California Institute of Technology (CIT)	California	$75,500.00	$123,000.00	NaN	$104,000.00	$161,000.00	NaN
2	Harvey Mudd College	California	$71,800.00	$122,000.00	NaN	$96,000.00	$180,000.00	NaN
3	University of California, Berkeley	California	$59,900.00	$112,000.00	$59,500.00	$81,000.00	$149,000.00	$201,000.00
4	Occidental College	California	$51,900.00	$105,000.00	NaN	$54,800.00	$157,000.00	NaN

In [7]:
```
region_df.shape
```

Out[7]: (320, 8)

Data Cleaning

Dropping Extraneous Columns

In [8]:
```
college_type_df = college_type_df.drop(['Mid-Career 10th Percentile Salary',
                                        'Mid-Career 25th Percentile Salary',
                                        'Mid-Career 75th Percentile Salary',
                                        'Mid-Career 90th Percentile Salary'], axis=1)
```

In [9]:
```
region_df = region_df.drop(['Mid-Career 10th Percentile Salary',
                            'Mid-Career 25th Percentile Salary',
                            'Mid-Career 75th Percentile Salary',
                            'Mid-Career 90th Percentile Salary'], axis=1)
```

Merging College Type and Region Data

In [10]:
```
df = college_type_df.merge(region_df, on=['School Name']).drop(['Starting Median Salary_y',
                                                                'Mid-Career Median Salary_y'], axis=1)
```

In [11]:
```
df = df.rename(columns={'Starting Median Salary_x': 'Starting Median Salary',
                        'Mid-Career Median Salary_x':  'Mid-Career Median Salary'})
```

Embry-Riddle Aeronautical University (ERAU) has no region, so it is dropped after merging.

In [12]:
```
pd.concat([college_type_df,df.drop(['Region'], axis=1)]).drop_duplicates(keep=False)
```

Out[12]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
17	Embry-Riddle Aeronautical University (ERAU)	Engineering	$52,700.00	$80,700.00

Cleaning Column Names

In [13]:
```
df = df.rename(columns={'School Name': 'school_name',
                        'School Type': 'school_type',
                        'Region': 'region',
                        'Starting Median Salary': 'starting_median_salary',
                        'Mid-Career Median Salary':  'mid_career_median_salary'})
```

In [14]:
```
df = df[['school_name', 'school_type', 'region', 'starting_median_salary', 'mid_career_median_salary']]
```

In [15]:
```
df.head()
```

Out[15]:

	school_name	school_type	region	starting_median_salary	mid_career_median_salary
0	Massachusetts Institute of Technology (MIT)	Engineering	Northeastern	$72,200.00	$126,000.00
1	California Institute of Technology (CIT)	Engineering	California	$75,500.00	$123,000.00
2	Harvey Mudd College	Engineering	California	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Engineering	Northeastern	$62,400.00	$114,000.00
4	Cooper Union	Engineering	Northeastern	$62,200.00	$114,000.00

Remove Party Schools

In [16]:
```
df = df[df.school_type != 'Party']
```

In [17]:
```
df.shape
```

Out[17]: (248, 5)

Rename Engineering Schools to Polytechnic

In [18]:
```
df['school_type'].replace({'Engineering': 'Polytechnic'}, inplace=True)
```

Remove Dollar Signs, Convert Salaries to Floats

In [19]:
```
df['starting_median_salary'] = df['starting_median_salary'].str.replace('$', '')
df['starting_median_salary'] = df['starting_median_salary'].str.replace(',', '')
df['starting_median_salary'] = df['starting_median_salary'].astype(float)

df['mid_career_median_salary'] = df['mid_career_median_salary'].str.replace('$', '')
df['mid_career_median_salary'] = df['mid_career_median_salary'].str.replace(',', '')
df['mid_career_median_salary'] = df['mid_career_median_salary'].astype(float)
```

Create Salary Percentage Increase Column

In [20]:
```
df['salary_percentage_increase'] = df.apply(lambda uni:
            (uni.mid_career_median_salary - uni.starting_median_salary) / uni.starting_median_salary, axis=1)
```

In [21]:
```
df.head()
```

Out[21]:

	school_name	school_type	region	starting_median_salary	mid_career_median_salary	salary_percentage_increase
0	Massachusetts Institute of Technology (MIT)	Polytechnic	Northeastern	72200.0	126000.0	0.745152
1	California Institute of Technology (CIT)	Polytechnic	California	75500.0	123000.0	0.629139
2	Harvey Mudd College	Polytechnic	California	71800.0	122000.0	0.699164
3	Polytechnic University of New York, Brooklyn	Polytechnic	Northeastern	62400.0	114000.0	0.826923
4	Cooper Union	Polytechnic	Northeastern	62200.0	114000.0	0.832797

In [22]:
```
df.shape
```

Out[22]:
```
(248, 6)
```

Descriptive Analysis

In [23]:
```
df.describe()
```

Out[23]:

	starting_median_salary	mid_career_median_salary	salary_percentage_increase
count	248.000000	248.000000	248.000000
mean	46070.161290	83884.677419	0.816592
std	6586.705784	14794.607533	0.147918
min	34800.000000	43900.000000	0.243626
25%	42000.000000	73400.000000	0.728739
50%	44750.000000	81550.000000	0.812348
75%	48150.000000	92850.000000	0.894470
max	75500.000000	134000.000000	1.310345

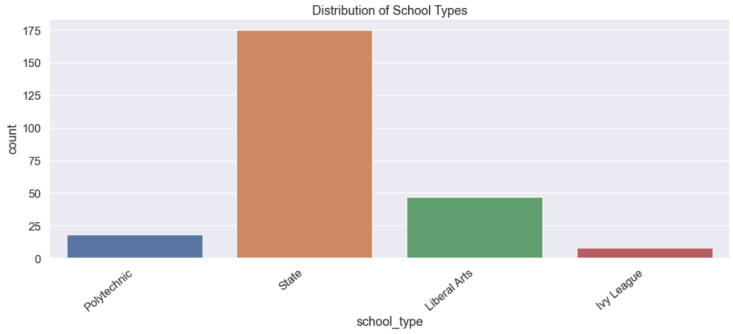
Distribution of School Types

In [24]:
```
df.groupby('school_type').count()
```

Out[24]:

school_type	school_name	region	starting_median_salary	mid_career_median_salary	salary_percentage_increase
Ivy League	8	8	8	8	8
Liberal Arts	47	47	47	47	47
Polytechnic	18	18	18	18	18
State	175	175	175	175	175

In [25]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Distribution of School Types')
ax = sns.countplot('school_type', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.tight_layout()
plt.show()
```



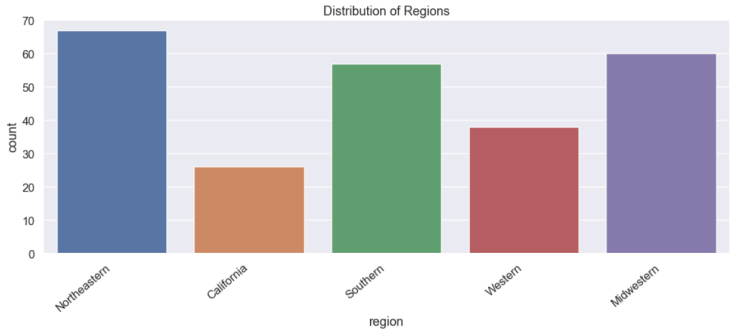
Distribution of Regions

In [26]:
```
df.groupby('region').count()
```

Out[26]:

region	school_name	school_type	starting_median_salary	mid_career_median_salary	salary_percentage_increase
California	26	26	26	26	26
Midwestern	60	60	60	60	60
Northeastern	67	67	67	67	67
Southern	57	57	57	57	57
Western	38	38	38	38	38

In [27]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Distribution of Regions')
ax = sns.countplot('region', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.tight_layout()
plt.show()
```



Which Universities Are Considered Western?

In [28]:
```
df[df['region'] == 'Western']
```

Out[28]:

	school_name	school_type	region	starting_median_salary	mid_career_median_salary	salary_percentage_increase
9	Colorado School of Mines	Polytechnic	Western	58100.0	106000.0	0.824441
15	New Mexico Institute of Mining and Technology ...	Polytechnic	Western	51000.0	93400.0	0.831373
37	Arizona State University (ASU)	State	Western	47400.0	84100.0	0.774262
89	University of Puget Sound	Liberal Arts	Western	46600.0	81500.0	0.748927
90	Colorado College (CC)	Liberal Arts	Western	38500.0	81400.0	1.114286
91	Reed College	Liberal Arts	Western	40500.0	81100.0	1.002469
93	Whitman College	Liberal Arts	Western	43500.0	80100.0	0.841379
100	Lewis & Clark College	Liberal Arts	Western	38900.0	72600.0	0.866324
101	Fort Lewis College	Liberal Arts	Western	42000.0	69800.0	0.661905
103	Evergreen State College	Liberal Arts	Western	39500.0	63900.0	0.617722
118	University of Colorado - Boulder (UCB)	State	Western	47100.0	97600.0	1.072187
136	University of Arizona	State	Western	47500.0	86100.0	0.812632
138	University of Washington (UW)	State	Western	48800.0	85300.0	0.747951
143	Washington State University (WSU)	State	Western	45300.0	84700.0	0.869757
147	University of Colorado - Denver	State	Western	46100.0	84400.0	0.830803
154	Oregon State University (OSU)	State	Western	45100.0	83300.0	0.847007
155	University of Utah	State	Western	45400.0	83200.0	0.832599
156	University of Nevada, Reno (UNR)	State	Western	46500.0	82900.0	0.782796
162	University of Idaho	State	Western	44900.0	82000.0	0.826281
166	University of New Mexico (UNM)	State	Western	41600.0	81600.0	0.961538
178	New Mexico State University	State	Western	44300.0	79500.0	0.794582
181	Colorado State University (CSU)	State	Western	44800.0	79000.0	0.763393
183	University of Wyoming (UW)	State	Western	44500.0	78700.0	0.768539
184	Utah State University	State	Western	43800.0	78700.0	0.796804
186	University of Oregon	State	Western	42200.0	78400.0	0.857820
193	Montana State University - Bozeman	State	Western	46600.0	77500.0	0.663090
198	University of Hawaii	State	Western	43800.0	76000.0	0.735160
202	Western Washington University	State	Western	42700.0	75400.0	0.765808
211	Idaho State University	State	Western	44900.0	73400.0	0.634744
214	University of Alaska, Anchorage	State	Western	45900.0	72600.0	0.581699
223	University of Montana	State	Western	37300.0	71900.0	0.927614
226	University of Nevada, Las Vegas (UNLV)	State	Western	45200.0	71600.0	0.584071
233	Portland State University (PSU)	State	Western	42600.0	70900.0	0.664319
234	Eastern Washington University	State	Western	38600.0	70900.0	0.836788
240	Boise State University (BSU)	State	Western	40800.0	69500.0	0.703431
247	Utah Valley State College	State	Western	42400.0	67100.0	0.582547
265	Southern Utah University	State	Western	41900.0	56500.0	0.348449
266	Montana State University - Billings	State	Western	37900.0	50600.0	0.335092

Exploratory Data Analysis

Correlation Between Features

In [29]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Correlation of Features')
ax = sns.heatmap(df.corr(), cmap="YlGnBu", annot=True)
```

Number of Schools With Salary Percentage Increase < 65% By School Type

In [30]:
```
df[df['salary_percentage_increase'] < 0.65].groupby('school_type').count()
```

Out[30]:

school_type	school_name	region	starting_median_salary	mid_career_median_salary	salary_percentage_increase
Liberal Arts	2	2	2	2	2
Polytechnic	1	1	1	1	1
State	21	21	21	21	21

Number of Schools With Salary Percentage Increase < 65% By Region

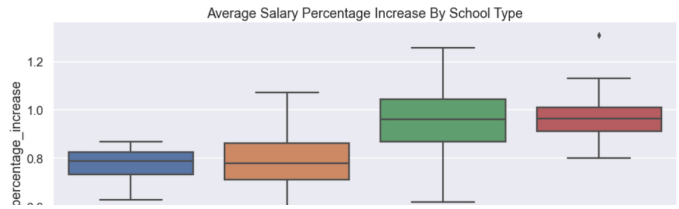
In [31]:
```
df[df['salary_percentage_increase'] < 0.65].groupby('region').count()
```

Out[31]:

region	school_name	school_type	starting_median_salary	mid_career_median_salary	salary_percentage_increase
California	2	2	2	2	2
Midwestern	9	9	9	9	9
Northeastern	2	2	2	2	2
Southern	4	4	4	4	4
Western	7	7	7	7	7

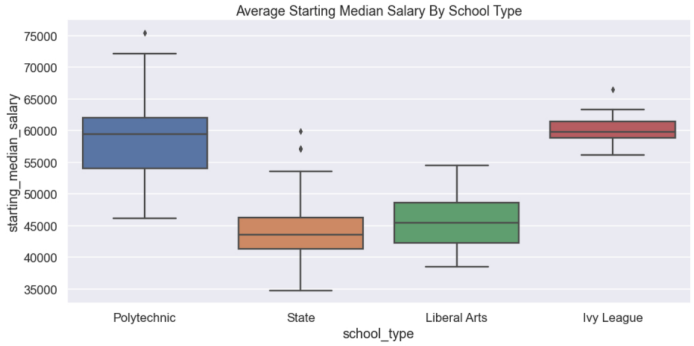
Average Salary Percentage Increase By School Type

In [32]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Salary Percentage Increase By School Type')
ax = sns.boxplot(x="school_type", y="salary_percentage_increase", data=df)
```



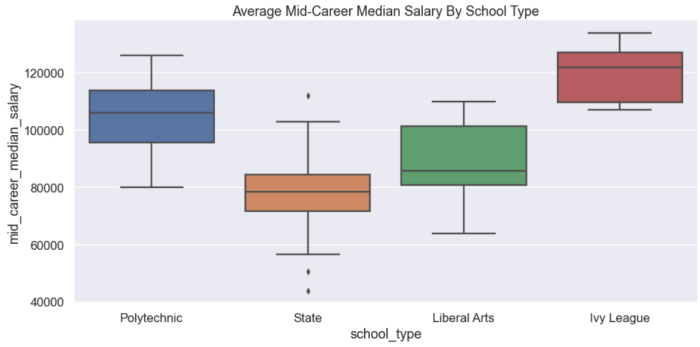
Average Starting Median Salary By School Type

In [33]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Starting Median Salary By School Type')
ax = sns.boxplot(x="school_type", y="starting_median_salary", data=df)
```



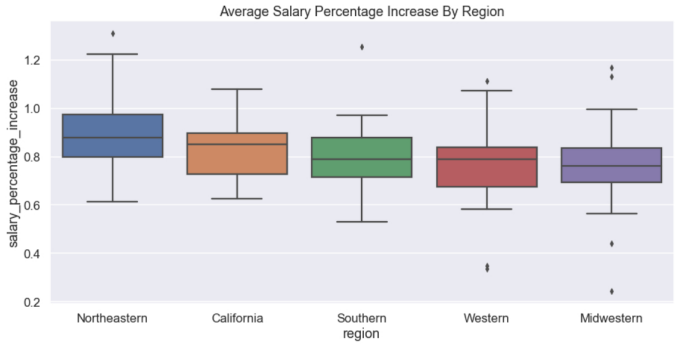
Average Mid-Career Median Salary By School Type

In [34]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Mid-Career Median Salary By School Type')
ax = sns.boxplot(x="school_type", y="mid_career_median_salary", data=df)
```



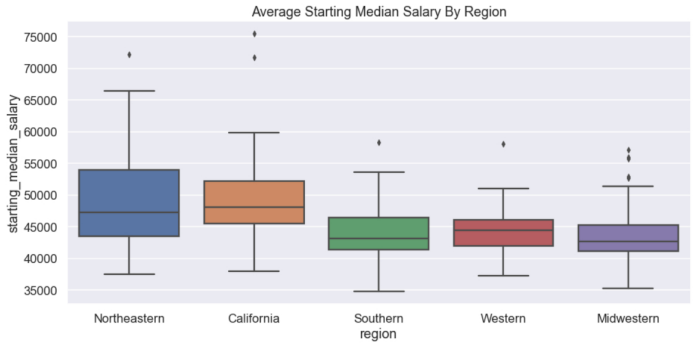
Average Salary Percentage Increase By Region

In [35]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Salary Percentage Increase By Region')
ax = sns.boxplot(x="region", y="salary_percentage_increase", data=df)
```



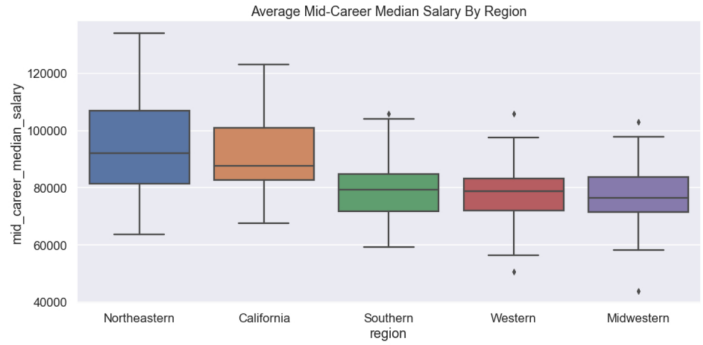
Average Starting Median Salary By Region

In [36]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Starting Median Salary By Region')
ax = sns.boxplot(x="region", y="starting_median_salary", data=df)
```



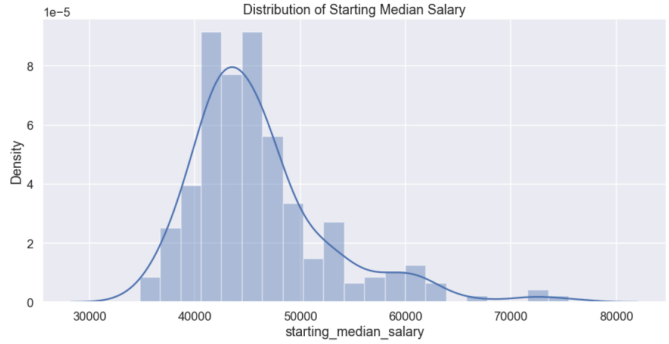
Average Mid-Career Median Salary By Region

In [37]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Mid-Career Median Salary By Region')
ax = sns.boxplot(x="region", y="mid_career_median_salary", data=df)
```

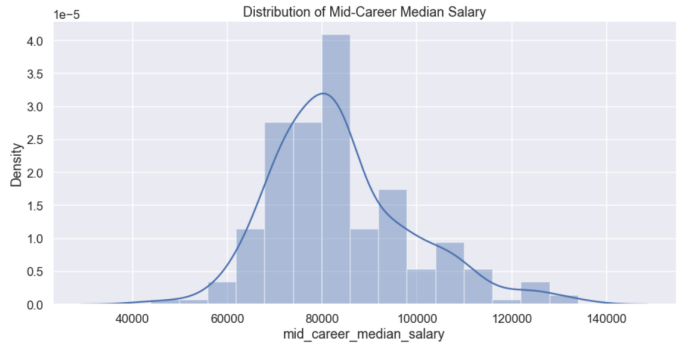


Distributions of Salary Data

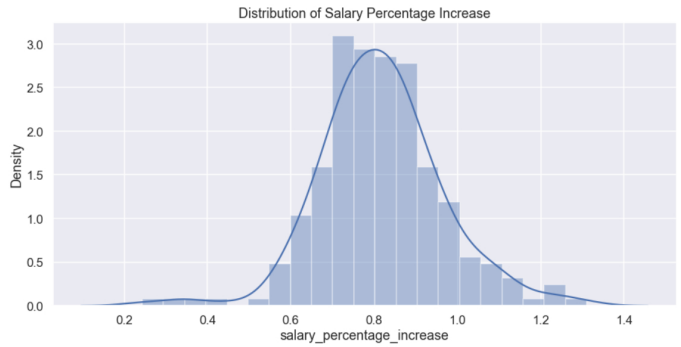
In [38]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Distribution of Starting Median Salary')
ax = sns.distplot(df['starting_median_salary'])
```



In [39]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Distribution of Mid-Career Median Salary')
ax = sns.distplot(df['mid_career_median_salary'])
```



In [40]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Distribution of Salary Percentage Increase')
ax = sns.distplot(df['salary_percentage_increase'])
```



T-Tests for School Types (Liberal Arts Control)

Salary Percentage Increase T-Test Between Liberal Arts and Polytechnic Schools

In [41]:
```
liberal_arts_salary_percentage_increase = df[df['school_type'] == 'Liberal Arts'].salary_percentage_increase
polytechnic_salary_percentage_increase = df[df['school_type'] == 'Polytechnic'].salary_percentage_increase
ttest_ind(liberal_arts_salary_percentage_increase, polytechnic_salary_percentage_increase)
```

Out[41]:
```
Ttest_indResult(statistic=4.800705454332984, pvalue=1.0115755771892046e-05)
```

Salary Percentage Increase T-Test Between Liberal Arts and Ivy League Schools

In [42]:
```
liberal_arts_salary_percentage_increase = df[df['school_type'] == 'Liberal Arts'].salary_percentage_increase
ivy_league_salary_percentage_increase = df[df['school_type'] == 'Ivy League'].salary_percentage_increase
ttest_ind(liberal_arts_salary_percentage_increase, ivy_league_salary_percentage_increase)
```

Out[42]:
```
Ttest_indResult(statistic=-0.6512981067651545, pvalue=0.5176682164702404)
```

Salary Percentage Increase T-Test Between Liberal Arts and State Schools

In [43]:
```
liberal_arts_salary_percentage_increase = df[df['school_type'] == 'Liberal Arts'].salary_percentage_increase
state_salary_percentage_increase = df[df['school_type'] == 'State'].salary_percentage_increase
ttest_ind(liberal_arts_salary_percentage_increase, state_salary_percentage_increase)
```

Out[43]: Ttest_indResult(statistic=8.003199455278951, pvalue=6.919197985464 79e-14)

Starting Median Salary T-Test Between Liberal Arts and Polytechnic Schools

In [44]:
```
liberal_arts_starting_salary = df[df['school_type'] == 'Liberal Arts'].starting_median_salary
polytechnic_starting_salary = df[df['school_type'] == 'Polytechnic'].starting_median_salary
ttest_ind(liberal_arts_starting_salary, polytechnic_starting_salary)
```

Out[44]: Ttest_indResult(statistic=-8.877728212474686, pvalue=1.0543462424301967e-12)

Starting Median Salary T-Test Between Liberal Arts and Ivy League Schools

In [45]:
```
liberal_arts_starting_salary = df[df['school_type'] == 'Liberal Arts'].starting_median_salary
ivy_league_starting_salary = df[df['school_type'] == 'Ivy League'].starting_median_salary
ttest_ind(liberal_arts_starting_salary, ivy_league_starting_salary)
```

Out[45]: Ttest_indResult(statistic=-9.093291799920882, pvalue=2.0957638591438893e-12)

Starting Median Salary T-Test Between Liberal Arts and State Schools

In [46]:
```
liberal_arts_starting_salary = df[df['school_type'] == 'Liberal Arts'].starting_median_salary
state_starting_salary = df[df['school_type'] == 'State'].starting_median_salary
ttest_ind(liberal_arts_starting_salary, state_starting_salary)
```

Out[46]: Ttest_indResult(statistic=2.299200067779707, pvalue=0.022431861093122757)

Mid-Career Median Salary T-Test Between Liberal Arts and Polytechnic Schools

In [47]:
```
liberal_arts_mid_career_salary = df[df['school_type'] == 'Liberal Arts'].mid_career_median_salary
polytechnic_mid_career_salary = df[df['school_type'] == 'Polytechnic'].mid_career_median_salary
ttest_ind(liberal_arts_mid_career_salary, polytechnic_mid_career_salary)
```

Out[47]: Ttest_indResult(statistic=-4.541018391196069, pvalue=2.5845163287204983e-05)

Mid-Career Median Salary T-Test Between Liberal Arts and Ivy League

In [48]:
```
liberal_arts_mid_career_salary = df[df['school_type'] == 'Liberal Arts'].mid_career_median_salary
ivy_league_mid_career_salary = df[df['school_type'] == 'Ivy League'].mid_career_median_salary
ttest_ind(liberal_arts_mid_career_salary, ivy_league_mid_career_salary)
```

Out[48]: Ttest_indResult(statistic=-6.6528297074972045, pvalue=1.632092083579218e-08)

Mid-Career Median Salary T-Test Between Liberal Arts and State Schools

In [49]:
```
liberal_arts_mid_career_salary = df[df['school_type'] == 'Liberal Arts'].mid_career_median_salary
state_mid_career_salary = df[df['school_type'] == 'State'].mid_career_median_salary
ttest_ind(liberal_arts_mid_career_salary, state_mid_career_salary)
```

Out[49]: Ttest_indResult(statistic=6.1213292585689585, pvalue=4.210926810452604e-09)

T-Tests for Regions (California Control)

Salary Percentage Increase T-Test Between California and Northeastern Schools

In [50]:
```
california_salary_percentage_increase = df[df['region'] == 'California'].salary_percentage_increase
northeastern_salary_percentage_increase = df[df['region'] == 'Northeastern'].salary_percentage_increase
ttest_ind(california_salary_percentage_increase, northeastern_salary_percentage_increase)
```

Out[50]: Ttest_indResult(statistic=-2.1921251305526286, pvalue=0.030922103016755 49)

Salary Percentage Increase T-Test Between California and Southern Schools

In [51]:
```
california_salary_percentage_increase = df[df['region'] == 'California'].salary_percentage_increase
southern_salary_percentage_increase = df[df['region'] == 'Southern'].salary_percentage_increase
ttest_ind(california_salary_percentage_increase, southern_salary_percentage_increase)
```

Out[51]: Ttest_indResult(statistic=1.016093181633718, pvalue=0.3126105052605623 3)

Salary Percentage Increase T-Test Between California and Western Schools

In [52]:
```
california_salary_percentage_increase = df[df['region'] == 'California'].salary_percentage_increase
western_salary_percentage_increase = df[df['region'] == 'Western'].salary_percentage_increase
ttest_ind(california_salary_percentage_increase, western_salary_percentage_increase)
```

Out[52]: Ttest_indResult(statistic=1.612641663303708, pvalue=0.1119015187166127 2)

Salary Percentage Increase T-Test Between California and Midwestern Schools

In [53]:
```
california_salary_percentage_increase = df[df['region'] == 'California'].salary_percentage_increase
midwestern_salary_percentage_increase = df[df['region'] == 'Midwestern'].salary_percentage_increase
ttest_ind(california_salary_percentage_increase, midwestern_salary_percentage_increase)
```

Out[53]: Ttest_indResult(statistic=1.898184112009617 2, pvalue=0.061106158986560466)

Starting Median Salary T-Test Between California and Northeastern Schools

In [54]:
```
california_starting_salary = df[df['region'] == 'California'].starting_median_salary
northeastern_starting_salary = df[df['region'] == 'Northeastern'].starting_median_salary
ttest_ind(california_starting_salary, northeastern_starting_salary)
```

Out[54]: Ttest_indResult(statistic=0.3779198066764735, pvalue=0.706370256665070 7)

Starting Median Salary T-Test Between California and Southern Schools

In [55]:
```
california_starting_salary = df[df['region'] == 'California'].starting_median_salary
southern_starting_salary = df[df['region'] == 'Southern'].starting_median_salary
ttest_ind(california_starting_salary, southern_starting_salary)
```

Out[55]: Ttest_indResult(statistic=4.228068774109023, pvalue=6.151185841076348e-05)

Starting Median Salary T-Test Between California and Western Schools

In [56]:
```
california_starting_salary = df[df['region'] == 'California'].starting_median_salary
western_starting_salary = df[df['region'] == 'Western'].starting_median_salary
ttest_ind(california_starting_salary, western_starting_salary)
```

Out[56]: Ttest_indResult(statistic=3.883443061979234, pvalue=0.000252639240359613 16)

Starting Median Salary T-Test Between California and Midwestern Schools

In [57]:
```
california_starting_salary = df[df['region'] == 'California'].starting_median_salary
midwestern_starting_salary = df[df['region'] == 'Midwestern'].starting_median_salary
ttest_ind(california_starting_salary, midwestern_starting_salary)
```

Out[57]: Ttest_indResult(statistic=4.65158037677967, pvalue=1.2145880312169153e-05)

Mid-Career Median Salary T-Test Between California and Northeastern Schools

In [58]:
```
california_mid_career_salary = df[df['region'] == 'California'].mid_career_median_salary
northeastern_mid_career_salary = df[df['region'] == 'Northeastern'].mid_career_median_salary
ttest_ind(california_mid_career_salary, northeastern_mid_career_salary)
```

Out[58]: Ttest_indResult(statistic=-0.6392830042608894, pvalue=0.524244845130057 1)

Mid-Career Median Salary T-Test Between California and Southern Schools

Mid-Career Median Salary T-Test Between California and Southern Schools

In [59]:
```
california_mid_career_salary = df[df['region'] == 'California'].mid_career_median_salary
southern_mid_career_salary = df[df['region'] == 'Southern'].mid_career_median_salary
ttest_ind(california_mid_career_salary, southern_mid_career_salary)
```

Out[59]: Ttest_indResult(statistic=4.07510052100244 65, pvalue=0.00010668633994892891)

Mid-Career Median Salary T-Test Between California and Western Schools

In [60]:
```
california_mid_career_salary = df[df['region'] == 'California'].mid_career_median_salary
western_mid_career_salary = df[df['region'] == 'Western'].mid_career_median_salary
ttest_ind(california_mid_career_salary, western_mid_career_salary)
```

Out[60]: Ttest_indResult(statistic=4.3936108385668575, pvalue=4.439192828355467e-05)

Mid-Career Median Salary T-Test Between California and Midwestern Schools

In [61]:
```
california_mid_career_salary = df[df['region'] == 'California'].mid_career_median_salary
midwestern_mid_career_salary = df[df['region'] == 'Midwestern'].mid_career_median_salary
ttest_ind(california_mid_career_salary, midwestern_mid_career_salary)
```

Out[61]: Ttest_indResult(statistic=5.121775665960744, pvalue=1.8968772842298055e-06)

Hypothesis Analysis

Function to Create Hypothesis Column (Distinguishes Between Hypothesis Yes/No)

In [62]:
```
def hypothesis_filter(uni):
    if (uni.school_type == 'Polytechnic' and uni.region == 'California') or (
        uni.school_type == 'Polytechnic' and uni.region == 'Western') or (
        uni.school_type == 'Polytechnic' and uni.region == 'Northeastern') or (
        uni.school_type == 'Ivy League'):
        return 'Yes Hypothesis'
    return 'No Hypothesis'
```

In [63]:
```
df['hypothesis'] = df.apply(hypothesis_filter, axis=1)
```

Comparing Hypothesis Yes/No Average Salary Percentage Increase

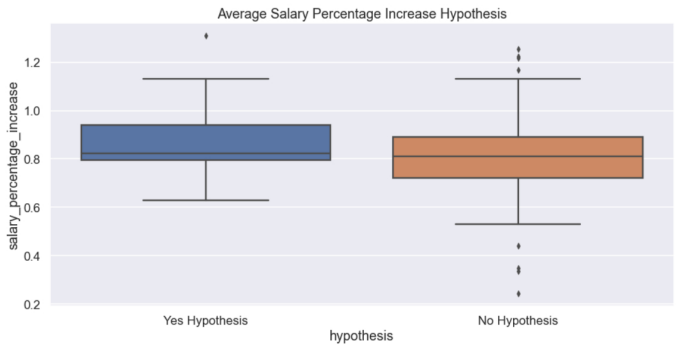
In [64]:
```
df[df['hypothesis'] == 'Yes Hypothesis'].salary_percentage_increase.mean()
```

Out[64]: 0.8592922933659525

In [65]:
```
df[df['hypothesis'] == 'No Hypothesis'].salary_percentage_increase.mean()
```

Out[65]: 0.8126416625562957

In [66]:
```
fig, ax = plt.subplots(figsize=(15,7))
ax.set_title('Average Salary Percentage Increase Hypothesis')
ax = sns.boxplot(x="hypothesis", y="salary_percentage_increase", data=df)
```



Salary Percentage Increase T-Test Between Hypothesis Yes/No

In [67]:
```
ttest_ind(df[df['hypothesis'] == 'Yes Hypothesis'].salary_percentage_increase,
          df[df['hypothesis'] == 'No Hypothesis'].salary_percentage_increase)
```

Out[67]: Ttest_indResult(statistic=1.3852904579934884, pvalue=0.16721848259285121)