What role does urbanization, using China as an example, play in shaping average temperature trends in a warming world?

Introduction

With the development of modern cities and urbanization, the global average temperature has been increasing over the past century, leading to concerns about the impacts of climate change. While urbanization significantly improves life standards and society efficiency, it is commonly believed to be a significant contributor to the global warming process.

China, one of the fastest-growing developing countries in the world, is undergoing rapid urbanization in the recent few decades. This research aims to examine the role of urbanization in shaping average temperature trends in China. The analysis will be focused on finding the relationship between average temperature and other independent variables such as city, years, and percentage change in annual temperature based on a time series analysis starting from 1950 to 2012. Specifically, 1979 to 2012 is the period China decided to begin the journey of reforming and opening up. We will use year as independent variable to show the trend of annual temperature change between 1950 to 2012. We will also study the difference in annual average temperature using city as independent variable. We will use percentage in annual average temperature as the independent variable to see the frequency of cities that has a temperature over certain value.

By investigating the impact of urbanization on temperature trends in China, this study will further enhance our understanding of the interaction between human activity and the environment. We find out that the average temperature is increasing overall between 1978 to 2012, and remained stable between 1950 to 1978. This shows that the relationship between urbanization and economic development is positively related to the warming of China. The overall average temperature is increasing in China, with more extreme high-temperature years and less low-temperature years.

Data Cleaning

Basic data cleaning

In [1]:
```
# Imports
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Read dataset
df = pd.read_csv('/Users/booker/Desktop/ECO225Project/Data/GlobalLandTemperat
```

In [2]:
```
# Check missing values
print(df.isnull().sum())

# Replacing missing values (not dropping since we will calculate the temperat
df['AverageTemperature'].interpolate(method='linear', inplace=True)
```

```
dt                                     0
AverageTemperature                364130
AverageTemperatureUncertainty     364130
City                                   0
Country                                0
Latitude                               0
Longitude                              0
dtype: int64
```

We will first convert date to year for a easier calculation and groupby in the following steps.

In [3]:
```
# Convert all dates
df['Date'] = pd.to_datetime(df.dt)
df.drop(columns = ['dt'], axis = 1, inplace = True)
df['Year'] = df['Date'].dt.year
```

Line Plots data cleaning

Group by China and year to get national average temperature, and the filter the year into three periods to create three datasets used to plot the line plots.

In [4]:
```
# Group by China and year to get national average temperature
df_china = df[df['Country'] == 'China']
grouped_na = df_china.groupby('Year').mean().reset_index()
grouped_na
```

Out[4]:

	Year	AverageTemperature	AverageTemperatureUncertainty
0	1816	12.992386	2.482814
1	1817	11.812655	2.168143
2	1818	12.491821	2.487060
3	1819	11.834655	2.122083
4	1820	7.468623	1.985787
...	...	...	...
193	2009	13.733718	0.301493
194	2010	13.430684	0.309203
195	2011	13.316098	0.334082
196	2012	13.204191	0.493267
197	2013	15.418062	0.627112

198 rows × 3 columns

In [5]:
```
# Filter the data to form 3 datasets ranging from different years
grouped_na_1 = grouped_na[(grouped_na['Year'] >= 1950) & (grouped_na['Year'] 
grouped_na_2 = grouped_na[(grouped_na['Year'] >= 1979) & (grouped_na['Year'] 
grouped_na_3 = grouped_na[(grouped_na['Year'] >= 1950) & (grouped_na['Year'] 
```

Bar Charts data cleaning

We first group data by city and year, and then find the top 10 cities with the largest and lowest temperature increased from 1950 to 2012. Then we use bar charts to plot the data.

In [6]:
```
# Select data from 1950 to 2012
grouped_time = df_china[(df_china['Year'] >= 1950) & (df_china['Year'] <= 201

# Group by City and Year to find the annual average temperature for each city
grouped_citi = grouped_time.groupby(['City', 'Year'])['AverageTemperature'].m
grouped_citi
```

Out[6]:

	City	Year	AverageTemperature
0	Acheng	1950	2.278167
1	Acheng	1951	2.380917
2	Acheng	1952	1.605833
3	Acheng	1953	1.878250
4	Acheng	1954	1.833333
...	...	...	...
23368	Ürümqi	2008	7.851250
23369	Ürümqi	2009	7.287417
23370	Ürümqi	2010	6.650083
23371	Ürümqi	2011	6.806083
23372	Ürümqi	2012	6.600167

23373 rows × 3 columns

In [7]:
```
# Find the top 10 cities with the largest temperature increased from 1950 to
grouped_citi.set_index(['City', 'Year'], inplace=True)

# Write the function that compute the percentage temperature change
def difference(df, first, last, column_name):
    for row in df.iterrows():
        index_value, columns_value = row
        start = df.loc[index_value[0], first]['AverageTemperature']
        end = df.loc[index_value[0], last]['AverageTemperature']
        cleaneddata = float((start - end)/end * 100)
        df.at[index_value, column_name] = cleaneddata

# Apply the function to compute the outcome
difference(grouped_citi, 2012, 1950, 'TemperaturePctChange')
```

In [8]:
```
# Find the top 10 cities with the largest temperature increased from 1950 to
largest_city = grouped_citi.sort_values(by='TemperaturePctChange', ascending=
top_cities = largest_city.reset_index().drop_duplicates(subset='City')
top_cities.set_index('Year', inplace=True)
top_cities_1950 = top_cities.loc[1950].head(10)
top_cities_1950
```

Out[8]:

Year	City	AverageTemperature	TemperaturePctChange
1950	Yakeshi	-1.619333	46.294772
1950	Shuangyashan	2.037833	36.967367
1950	Xining	2.497417	14.248056
1950	Qitaihe	3.010333	13.484110
1950	Didao	3.010333	13.484110
1950	Lianran	14.700083	5.242034
1950	Xianyang	11.395167	4.414152
1950	Xuchang	13.791750	4.322632
1950	Xingtai	11.706667	3.638952
1950	Tianjin	11.903917	3.267132

In [9]:

```
# Find the top 10 cities with the lowest temperature increased from 1950 to 2
lowest_city = grouped_citi.sort_values(by='TemperaturePctChange', ascending=T
low_cities = lowest_city.reset_index().drop_duplicates(subset='City')
low_cities.set_index('Year', inplace=True)
low_cities_1950 = low_cities.loc[1950].head(10)
low_cities_1950
```

Out[9]:

Year	City	AverageTemperature	TemperaturePctChange
1950	Dunhua	3.058000	-4.608132
1950	Yanji	3.058000	-4.608132
1950	Badaojiang	3.976250	-4.327780
1950	Gongzhuling	5.397167	-3.245530
1950	Ranghulu	3.322917	-3.134796
1950	Leiyang	17.465833	-2.219572
1950	Hengyang	17.465833	-2.219572
1950	Guilin	19.041083	-2.073149
1950	Kaili	15.173750	-1.754126
1950	Wencheng	11.765833	-1.531978

Pivot table

Create a pivot table to plot scatter plots for the top 4 cities with the largest and lowest temperature increased from 1950 to 2012. In addition, we can use bar charts to show the trend of the number of cities that exceeds 23 degrees and below 0 degrees under a 5 years interval.

In [10]:

```
# Create a pivot table
city_pivot = grouped_citi.pivot_table(values='AverageTemperature', index='Yea
city_pivot
```

Out[10]:

City	Acheng	Aksu	Altay	Anbu	Anda	Ankang	Anqing	Anqiu
Year								
1950	2.278167	9.318167	-0.995500	21.786750	4.079083	12.152917	15.907833	12.629583
1951	2.380917	9.732917	-0.021083	21.551583	4.090750	12.288417	15.818417	12.539417
1952	1.605833	9.559083	-1.072667	22.087833	3.308083	12.259083	15.522583	12.166667
1953	1.878250	10.244583	0.501500	22.031583	3.565667	12.844250	16.513667	12.842000
1954	1.833333	8.868000	-1.304417	21.897000	3.463417	12.023333	15.219833	11.859583
...	...	...	...	...	...	...	...	...
2008	4.137167	10.947167	1.910750	21.769083	5.843333	12.822500	16.315583	13.072500
2009	2.533333	11.262667	1.096583	22.193500	4.319500	13.080250	16.523833	13.266667
2010	2.337083	10.965667	0.060417	21.837417	3.852250	13.039083	16.440250	12.744333
2011	2.984833	10.674250	0.771417	21.586500	4.600250	12.776500	16.167417	12.717500
2012	2.522250	10.023333	0.483083	21.877167	4.071500	12.718500	16.151833	12.733000

63 rows × 371 columns

In [38]:

```
# Create a column that calculates the annual percentage change from 1950 to 2
mean_pivot = city_pivot.reset_index()
mean_pivot["mean_temp_change"] = mean_pivot.mean(axis=1)

# Filter the data from 1950 to 1978
mean_pivot_before = mean_pivot[mean_pivot['Year'] >= 1950]
mean_pivot_before = mean_pivot_before[mean_pivot_before['Year'] <= 1978]

# Filter the data from 1979 to 2012
mean_pivot_after = mean_pivot[mean_pivot['Year'] >= 1979]
mean_pivot_after = mean_pivot_after[mean_pivot_after['Year'] <= 2012]
```

Summary Statistics Tables

Table 1: Average temperature from 1950 to 1978

In this table, we can see that the average temperature from 1950 to 1978 is around 12.94 degrees, with a standard deviation of 0.32 degrees. The maximum temperature is 13.54 degrees, and the minimum temperature is 12.27 degrees. We can then compare the average temperature, maximum and minimum temperature of the temperature from 1979 to 2012 to see if there's a warming China.

In [11]:

```
# Describing the average temperature and average temperature uncertainty from
grouped_na_1.describe()
```

Out[11]:

	Year	AverageTemperature	AverageTemperatureUncertainty
count	29.000000	29.000000	29.000000
mean	1964.000000	12.942087	0.342187
std	8.514693	0.319699	0.053785
min	1950.000000	12.270832	0.278007

	Year	AverageTemperature	AverageTemperatureUncertainty
25%	1957.000000	12.783733	0.310153
50%	1964.000000	12.979855	0.325047
75%	1971.000000	13.183739	0.353979
max	1978.000000	13.544514	0.493532

Table 2: Average temperature from 1950 to 1978

In this table, we can see that the average temperature from 1979 to 2012 is around 13.42 degrees, with a standard deviation of 0.42 degrees. The maximum temperature is 14.26 degrees, and the minimum temperature is 12.50 degrees. All of the indicators suggest that there is a increasing temperature trend compared to the period of 1950 to 1978. It provides us some insights of the big picture before we plot our graphs.

In [12]:
```
# Describing the average temperature and average temperature uncertainty from
grouped_na_2.describe()
```

Out[12]:

	Year	AverageTemperature	AverageTemperatureUncertainty
count	34.000000	34.000000	34.000000
mean	1995.500000	13.415004	0.298044
std	9.958246	0.421474	0.037514
min	1979.000000	12.499564	0.262996
25%	1987.250000	13.183835	0.283851
50%	1995.500000	13.391446	0.289897
75%	2003.750000	13.722694	0.301469
max	2012.000000	14.267247	0.493267

Table 3: Average temperature from 1950 to 2012

In this table, we can see that the average temperature from 1950 to 2012 is around 13.20 degrees, with a standard deviation of 0.44 degrees. The maximum temperature is 14.26 degrees, and the minimum temperature is 12.27 degrees.

In [13]:
```
# Describing the average temperature and average temperature uncertainty from
grouped_na_3.describe()
```

Out[13]:

	Year	AverageTemperature	AverageTemperatureUncertainty
count	63.000000	63.000000	63.000000
mean	1981.000000	13.197312	0.318364
std	18.330303	0.444033	0.050471
min	1950.000000	12.270832	0.262996
25%	1965.500000	12.968142	0.288589
50%	1981.000000	13.187257	0.302814

	Year	AverageTemperature	AverageTemperatureUncertainty
75%	1996.500000	13.424886	0.327684
max	2012.000000	14.267247	0.493532

Table 4: Top 10 cities with the largest percent change in average temperature

We can see the top 10 cities with the largest percent change in average temperature from 1950 to 2012. The maximum annual temperature is 14.70 degrees among those 10 cities, while the lowest annual temperature is -1.62 degrees, with a very high standard deviation of 5.98 degrees. The city with the maximum percent change in annual temperature is 46.29%, showing that there is a huge variation in temperature for that city. It can be the result from global warming and urbanization.

In [14]:
```
top_cities_1950.describe()
```

Out[14]:

	AverageTemperature	TemperaturePctChange
count	10.000000	10.000000
mean	7.243417	14.536332
std	5.975593	15.103190
min	-1.619333	3.267132
25%	2.625646	4.345512
50%	7.202750	9.363072
75%	11.854604	14.057070
max	14.700083	46.294772

Table 5: Top 10 cities with the lowest percent change in average temperature

We can see the top 10 cities with the lowest percent change in average temperature from 1950 to 2012. The city with the minimum percent change in annual temperature is -4.61%, showing that there is a decrease in temperature for that city. We can also compare the previous graph and conclude that cities are becoming warmer.

In [15]:
```
low_cities_1950.describe()
```

Out[15]:

	AverageTemperature	TemperaturePctChange
count	10.000000	10.000000
mean	9.972467	-2.972277
std	6.843769	1.192123
min	3.058000	-4.608132
25%	3.486250	-4.057217
50%	8.581500	-2.677184

	AverageTemperature	TemperaturePctChange
75%	16.892812	-2.109755
max	19.041083	-1.531978

Plots, Histograms, Figures

Line Plots

In this section, we will first use line plots to demonstrate how the average temperature changed from 1950 to 2013 nationally.

Specifically, we will divide the time interval into two periods: 1950 to 1978 and 1979 to 2012. The first period is the founding of the People's Republic of China, whereas the second period is China decided to begin the journey of reforming and opening up. The second period represents the urbanization and fast economic development of China. We will also see the overall picture from 1950 to 2012.

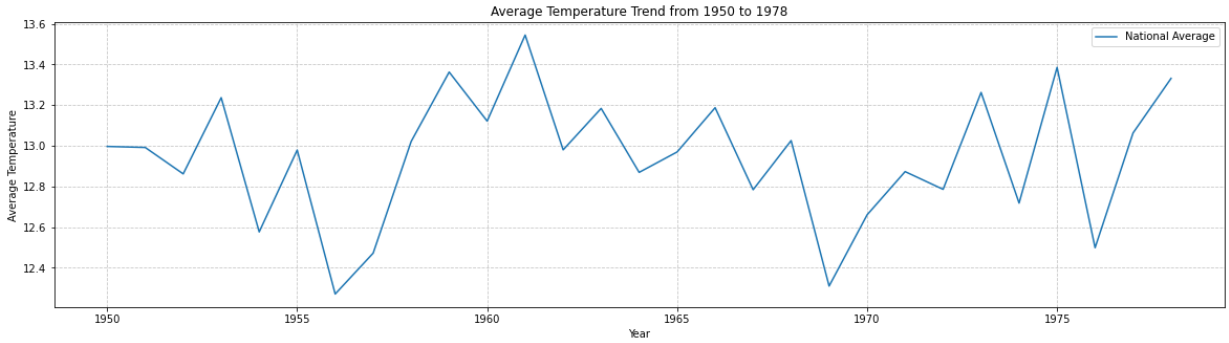
From the first graph, titled 'Average Temperature Trend from 1950 to 1978', we can see that the average temperature trend in China is relatively steady and is moving around at 13.0 degrees. This is the period of Chairman Mao Zedong's era, named after the founder of the Communist Party of China. Economic development was unstable and slow due to some policies that were originally aimed to achieve the goal of industrialization. The process of urbanization and modernization was slow.

From the second graph, 'Average Temperature Trend from 1979 to 2012', we can see that the average temperature trend in China is increasing. This is the period of the Deng Xiaoping era, named after the second leader that achieved economic recovery and development. He proposed a series of policies that aimed to accelerate the process of industrialization, including one of the most important policies in Chinese history: the Chinese reform.

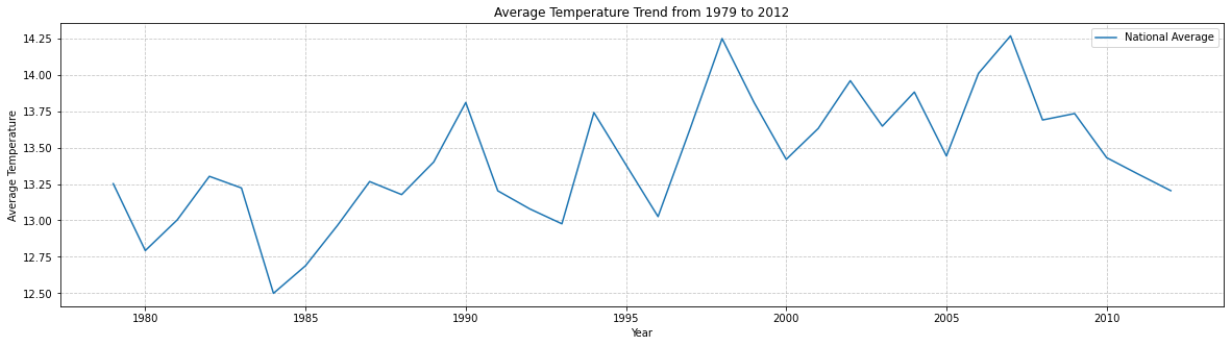
From the third graph, 'Average Temperature Trend from 1950 to 2012', we can see that the average temperature trend in China is increasing. This graph provides us with a clear image of how the average temperature change in China from 1950 to 2012. The temperature is positively correlated with the development of China, including economic development and urbanization.

In [61]:
```
# Write function that can generate line plots
def plot_lineplot(dataset, title):
    plt.figure(figsize=(20, 5))
    plt.plot(dataset['Year'], dataset['AverageTemperature'], label='National
    plt.xlabel('Year')
    plt.ylabel('Average Temperature')
    plt.title(title)
    plt.grid(linestyle='--', alpha=0.7)
    plt.legend()
    plt.show()
```

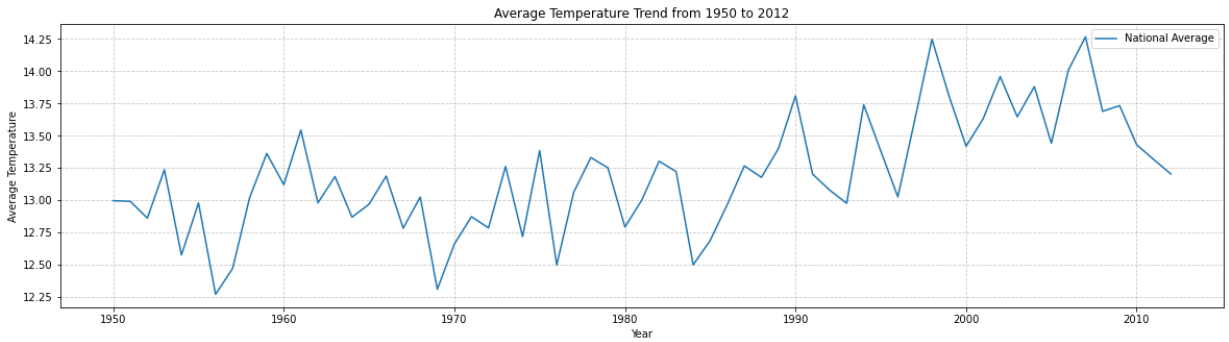
In [62]:
```
# Plot the graphs
plot_lineplot(grouped_na_1, 'Average Temperature Trend from 1950 to 1978')
```



In [63]:
```
plot_lineplot(grouped_na_2, 'Average Temperature Trend from 1979 to 2012')
```



In [64]:
```
plot_lineplot(grouped_na_3, 'Average Temperature Trend from 1950 to 2012')
```



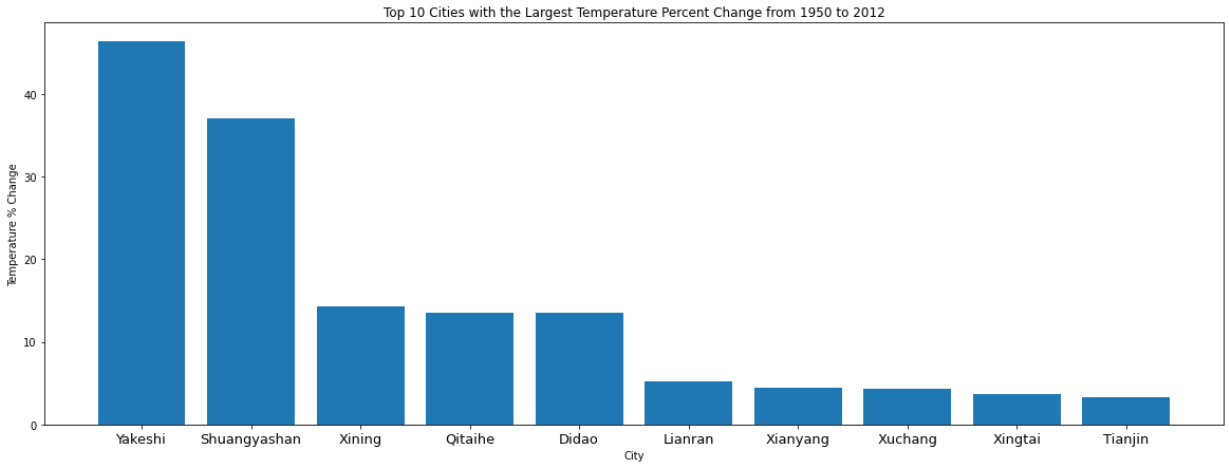
Bar charts

In this section, we will use bar charts to demonstrate the top 10 largest average temperature percent changes and the top 10 lowest average temperature percent changes from 1950 to 2012.

We can see that cities like Yakeshi and Shuangyashan experienced over 35 percent increase in annual temperature from 1950 to 2012, whereas the remaining cities increased around 10 percent in annual temperature. In contrast, Dunhua and Yanji experienced the lowest percentage change in temperature, but only up to around -4.5%. We can see that cities in China are becoming warmer based on the graphs below.

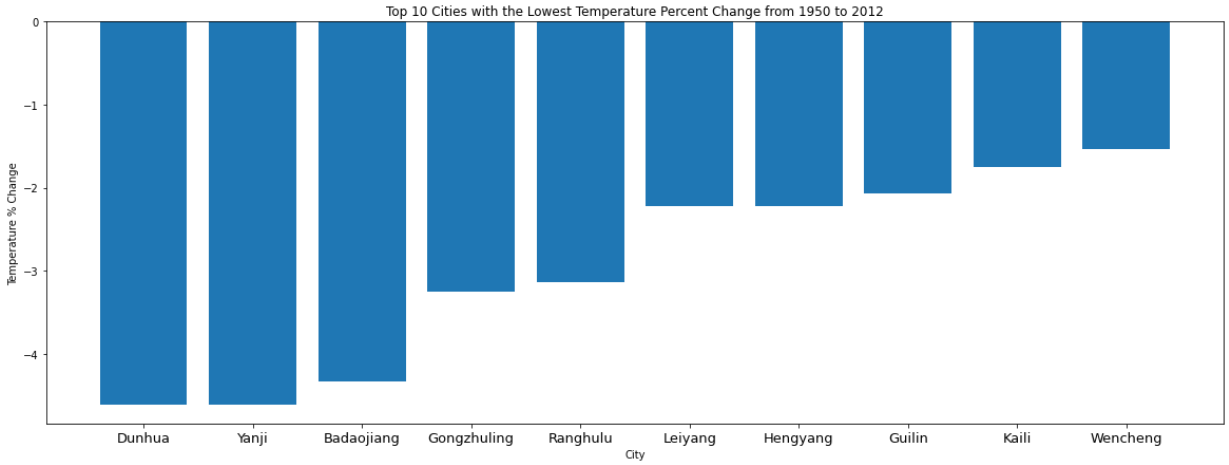
In [60]:
```
# Plot the graph of largest temperature change between 1950 to 2012
plt.figure(figsize=(20, 7))
plt.bar(top_cities_1950['City'], top_cities_1950['TemperaturePctChange'])
plt.xlabel('City')
plt.ylabel('Temperature % Change')
plt.title('Top 10 Cities with the Largest Temperature Percent Change from 195
```

```
plt.tick_params(axis='x', labelsize=13)
```



In [59]:

```
# Plot the graph of lowest temperature change between 1950 to 2012
plt.figure(figsize=(20, 7))
plt.bar(low_cities_1950['City'], low_cities_1950['TemperaturePctChange'])
plt.xlabel('City')
plt.ylabel('Temperature % Change')
plt.title('Top 10 Cities with the Lowest Temperature Percent Change from 1950
plt.tick_params(axis='x', labelsize=13)
plt.show()
```



We can see the trend of the number of cities that exceeds 23 degrees and below 0 degrees under a 5 years interval.

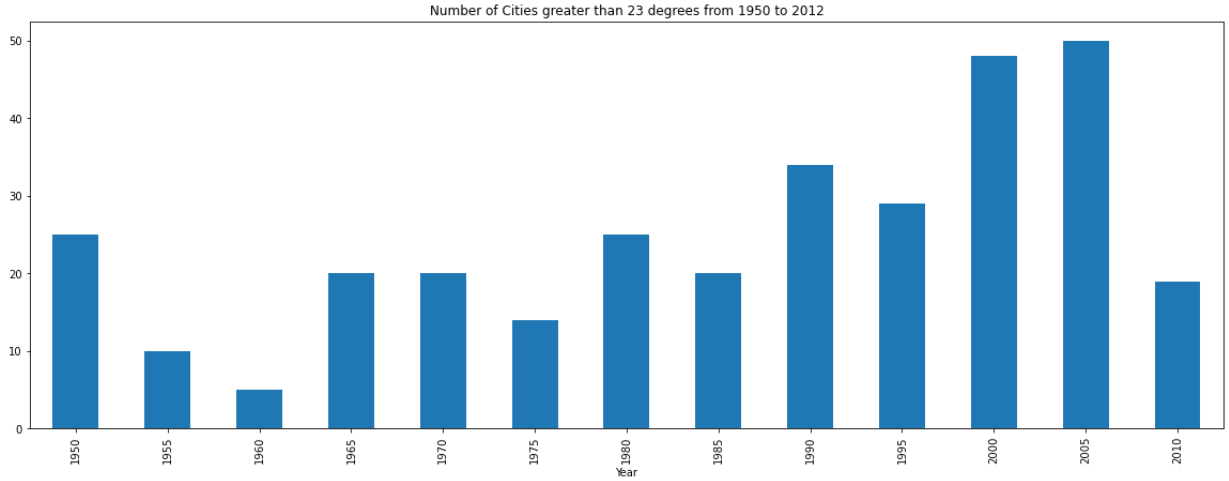
In the first graph, we see that there is an increasing trend from 1950 to 2010, which means that there are more and more cities that have an annual temperature of 23 degrees or more, especially for years over 1980. This can further prove that after 1980 when the urbanization and economic surge happened in China, the temperature is increasing faster.

In the second graph, we see that there is a decreasing trend from 1950 to 2010, which means that there are fewer cities that have an annual temperature of 0 degrees or less, meaning that all cities are becoming warmer, not colder. This can further prove that after 1980 when the urbanization and economic surge happened in China, cities are becoming hotter.

In [58]:

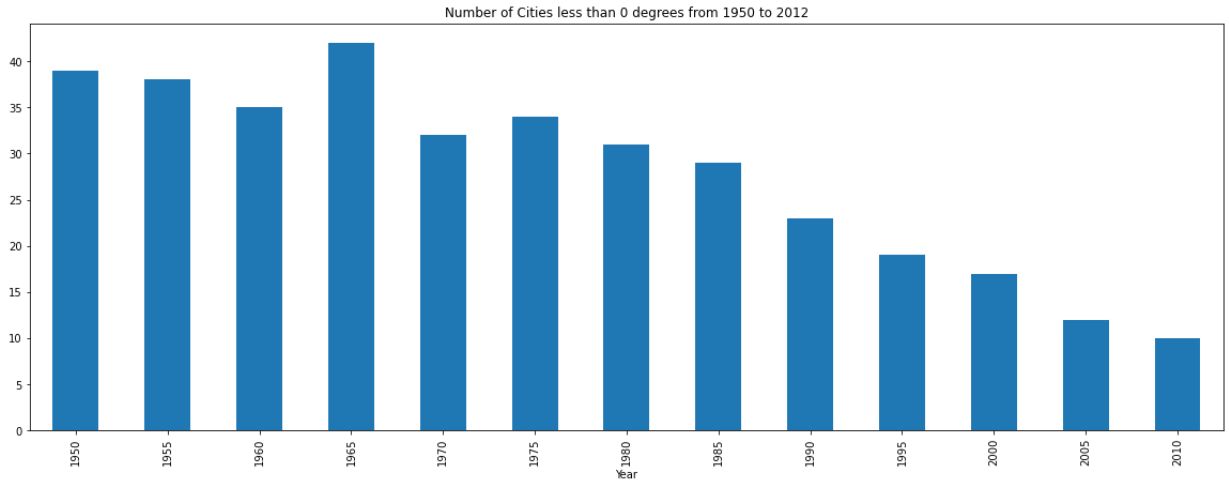
```
# Count the number of cities that is greater than 23 degrees
filt_temp = city_pivot[city_pivot > 23]
filt_count = filt_temp.count(axis=1)
grouped_count = filt_count.groupby(lambda x: x // 5 * 5).sum()
```

```
plt.figure(figsize=(20,7))
grouped_count.plot(kind='bar')
plt.title('Number of Cities greater than 23 degrees from 1950 to 2012')

plt.show()
```



In [57]:

```
# Count the number of cities that is lower than 0 degrees
filt_temp = city_pivot[city_pivot < 0]
filt_count = filt_temp.count(axis=1)
grouped_count = filt_count.groupby(lambda x: x // 5 * 5).sum()
plt.figure(figsize=(20,7))
grouped_count.plot(kind='bar')
plt.title('Number of Cities less than 0 degrees from 1950 to 2012')

plt.show()
```

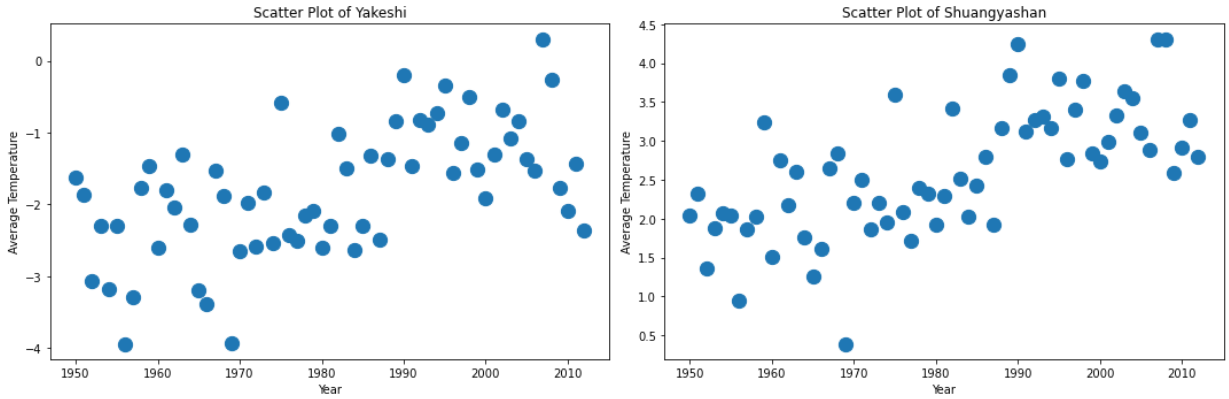


Scatter Plots

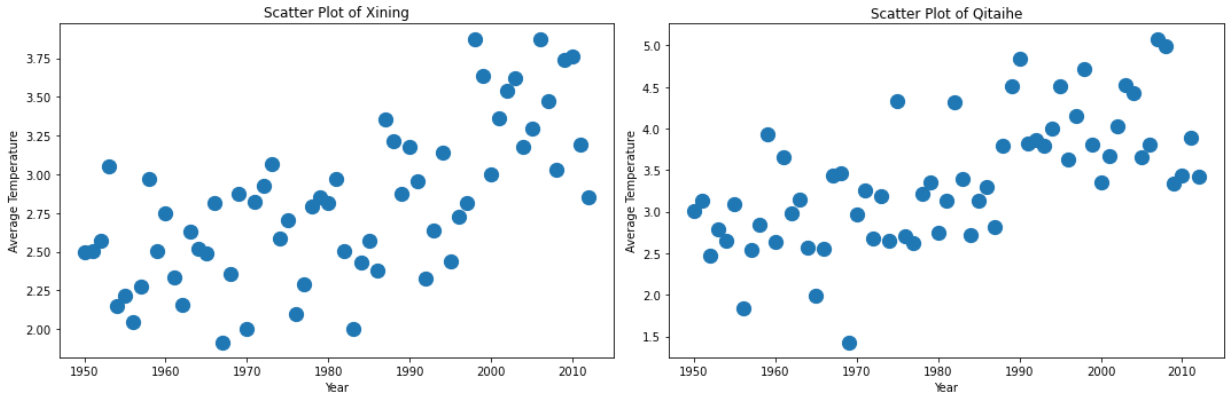
We want to have a closer look at the cities with the largest percentage change in average temperature.

From the four graphs below, which show the top 4 cities with the largest percentage change in average temperature, there is a strong positive relationship between time and annual temperature change. This result is not surprising because we already know that the average annual temperature is increasing overall in China from 1950 to 2012, but it gives us a more clear visualization of how the annual temperature change for the representative cities.

In [24]:
```
# Write a function to plot scatter plots
dff = city_pivot.reset_index()
def plot_scatter(city, title, ax):
    dff.plot(x='Year', y=city, kind='scatter', s=150, figsize=(15, 5), ax=ax)
    ax.set_title(title)
    ax.set_xlabel('Year')
    ax.set_ylabel('Average Temperature')

fig, ax = plt.subplots(1, 2, figsize=(10, 5))
plot_scatter('Yakeshi', 'Scatter Plot of Yakeshi', ax[0])
plot_scatter('Shuangyashan', 'Scatter Plot of Shuangyashan', ax[1])

plt.tight_layout()
```



In [25]:
```
fig, ax = plt.subplots(1, 2, figsize=(10, 5))
plot_scatter('Xining', 'Scatter Plot of Xining', ax[0])
plot_scatter('Qitaihe', 'Scatter Plot of Qitaihe', ax[1])
plt.tight_layout()
```



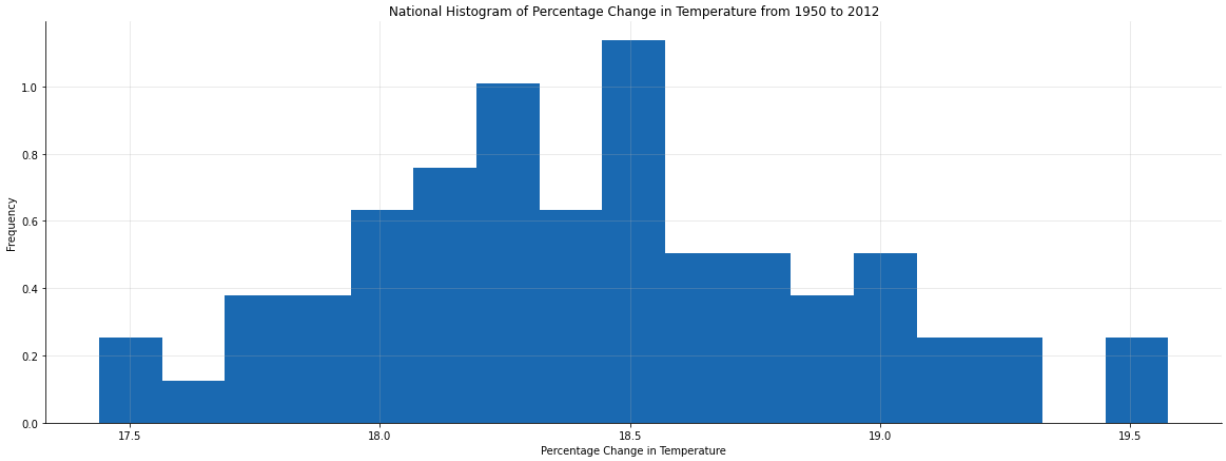
Histogram

We show a histogram of the percentage change in annual temperature over the period 1950 - 2012. We can observe that most of the cities have a percentage change of around 18.5. This shows that the majority of the city is experiencing a positive temperature change from 1950 to 2012. The overall shape of the histogram is approximately normal.

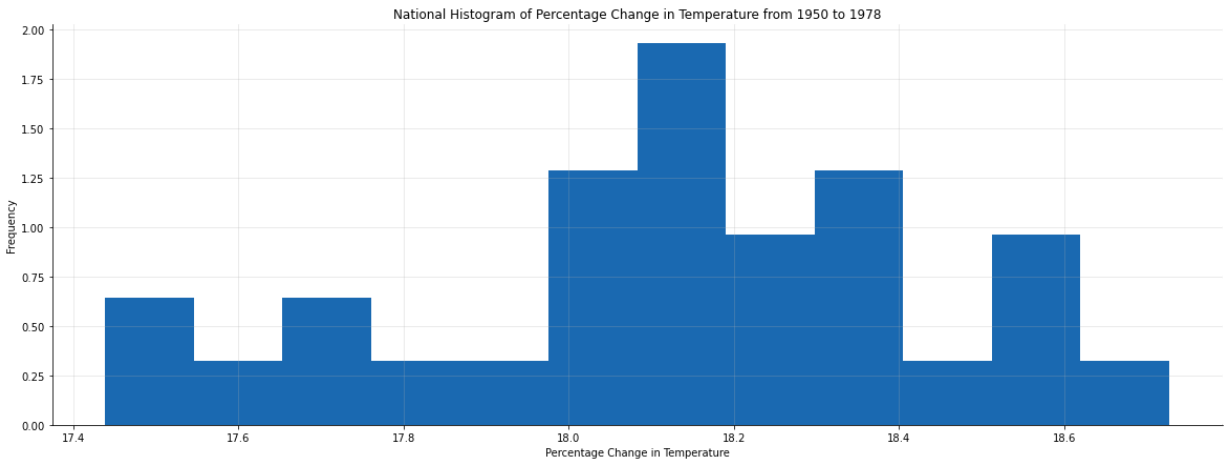
From 1950 to 1978, most cities have an annual percentage change in temperature of 18.1. However, from 1950 to 1978, most cities have an annual percentage change in temperature ranging from 18.25 to 19.00. This shows that the temperature is higher from 1979 to 2012, which is evidence proving that urbanization increases the temperature in China. Although

urbanization is beneficial to the human race and improves life standards and quality, it comes with the price of global warming.

In [56]:
```
fig, ax = plt.subplots(figsize=(20, 7))
mean_pivot.plot(
    kind="hist", y="mean_temp_change", color="#1a69b1",
    bins=17, legend=False, density=True, ax=ax
)
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
ax.set_xlabel("Percentage Change in Temperature")
ax.set_ylabel("Frequency")
ax.set_title("National Histogram of Percentage Change in Temperature from 195
ax.grid(True, alpha=0.3)
```

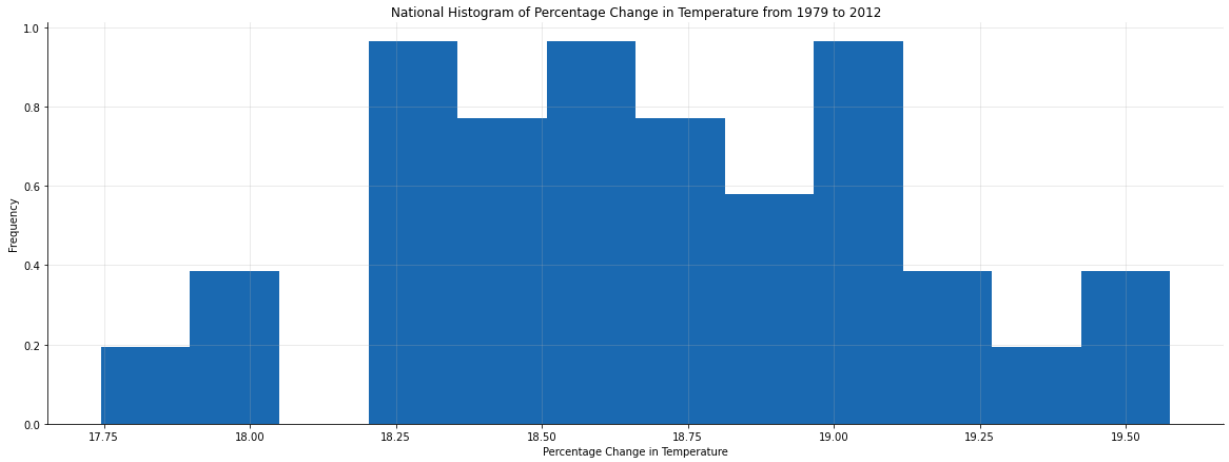


In [55]:
```
fig, ax = plt.subplots(figsize=(20, 7))
mean_pivot_before.plot(
    kind="hist", y="mean_temp_change", color="#1a69b1",
    bins=12, legend=False, density=True, ax=ax
)
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
ax.set_xlabel("Percentage Change in Temperature")
ax.set_ylabel("Frequency")
ax.set_title("National Histogram of Percentage Change in Temperature from 195
ax.grid(True, alpha=0.3)
```



In [54]:
```
fig, ax = plt.subplots(figsize=(20, 7))
mean_pivot_after.plot(
    kind="hist", y="mean_temp_change", color="#1a69b1",
```

```
    bins=12, legend=False, density=True, ax=ax
)
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
ax.set_xlabel("Percentage Change in Temperature")
ax.set_ylabel("Frequency")
ax.set_title("National Histogram of Percentage Change in Temperature from 197
ax.grid(True, alpha=0.3)
```



Conclusion

In conclusion, the research found that urbanization plays an important role in shaping average temperature trends in China from 1950 to 2012. As the graphs and evidence suggest, we see that the temperature is increasing at a higher speed from 1979 to 2012 (when urbanization happened and the economy starts to bloom in China) compared with 1950 to 1978 (when the economy is growing very slowly). In addition, the average annual temperature is increasing in China from 1950 to 2012. More and more cities experienced extremely high-temperature years and fewer cities experienced low-temperature years.

As urban areas start to grow, cities will experience higher temperatures due to potential factors such as industrialization and pollution, which can be discussed in future research. We can combine it with the population dataset for future implementation and study the relationship between population and temperature change. These findings are important to our understanding of global warming and may give us some suggestions for how to maintain the balance between society's development and the stability of the enviroment.