Report

Step 1: Import packages Create functions to load data, generate descriptive statistics for data, and generate summary statistics for data.

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt

def load_data(url):
    data = pd.read_csv(url)
    return data

def generate_descriptive_statistics(data):
    des_stat = data.describe()
    return des_stat


def generate_summary_statistics(filter_data):
    summary_stat = {
        "mean": filter_data.mean(),
        "median": filter_data.median(),
        "std": filter_data.std(),
    }
    return summary_stat
```

Step 2: Load data Filter data into Dataframe format and own the values we want.

In [3]:
```
url = "https://media.githubusercontent.com/media/nickeubank/MIDS_Data/master/World_Development_Indicators/wdi_small_tidy_2015.csv"
df = load_data(url)
df_filtered = df[
    [
        "Mortality rate, infant (per 1,000 live births)",
        "GDP per capita (constant 2010 US$)",
    ]
]
df_filtered
```

Out[3]:

	Mortality rate, infant (per 1,000 live births)	GDP per capita (constant 2010 US$)
0	54.9	574.184114
1	8.3	4524.684565
2	21.4	4776.787543
3	NaN	9835.410319
4	3.4	41767.526508
...	...	...
212	NaN	29408.055125
213	18.6	2631.811985
214	43.2	908.757846
215	44.2	1641.005482
216	40.2	1234.103352

217 rows × 2 columns

Using Pandas for descriptive statistics

In [4]:
```
des_stat = generate_descriptive_statistics(df_filtered)
print("Descriptive Statistics:\n", des_stat)
```

```
Descriptive Statistics:
       Mortality rate, infant (per 1,000 live births)  \
count                                      193.000000
mean                                        23.401036
std                                         21.063532
min                                          1.700000
25%                                          6.500000
50%                                         15.500000
75%                                         35.100000
max                                         91.600000

       GDP per capita (constant 2010 US$)
count                          198.000000
mean                         15335.724729
std                          22881.307340
min                            228.432544
25%                           1844.387439
50%                           6134.939066
75%                          17654.996438
max                         189464.583635
```

Generate summary statistics (mean, median, standard deviation)

In [5]:
```
summary_stat = generate_summary_statistics(df_filtered)
print("\nSummary Statistics:\n", summary_stat)
```

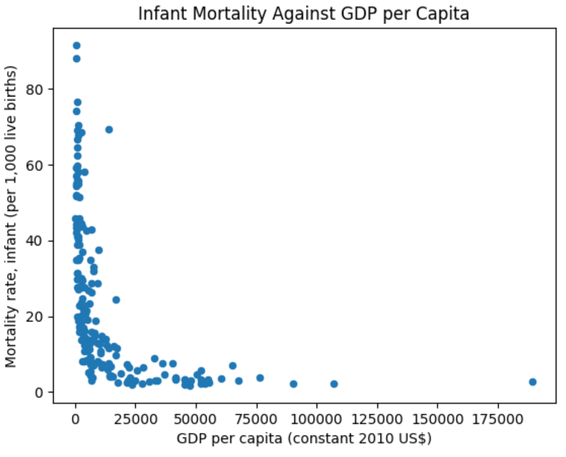
```
Summary Statistics:
 {'mean': Mortality rate, infant (per 1,000 live births)       23.401036
GDP per capita (constant 2010 US$)               15335.724729
dtype: float64, 'median': Mortality rate, infant (per 1,000 live births)      15.500000
GDP per capita (constant 2010 US$)               6134.939066
dtype: float64, 'std': Mortality rate, infant (per 1,000 live births)       21.063532
GDP per capita (constant 2010 US$)               22881.307340
dtype: float64}
```

Create one data visualization:

the relationship between GDP per capita (constant 2010 US$) and Mortality rate, infant (per 1,000 live births):

In [6]:
```
df_filtered.plot.scatter(
        x="GDP per capita (constant 2010 US$)",
        y="Mortality rate, infant (per 1,000 live births)",
        title="Infant Mortality Against GDP per Capita",
    )
print("\nVisualization:\n")
plt.show()
```

```
Visualization:
```



The relationship between "GDP per capita (constant 2010 US$)" and "Mortality rate, infant (per 1,000 live births)" is non-linear. We can find that some countries have significantly lower mortality rates than other.