# PRODIGY_DS_01

## Task 1. To create a bar chart or Histogram for a continuous or categorical variable.

INTRODUCTION : We will analyse the data set via curating different bar charts and histogram.

DATA DESCRIPTION: We have taken a data set of different country with their annual working hours for different years

### STEP1. Loading data set

In [2]:
```
import pandas as pd
Working_hours_data=pd.read_csv("C:/Users/Ujjaval Raj/Downloads/annual-working-hours
```

### STEP2. Understanding data set

In [3]:
```
print(Working_hours_data.head)
```

```
<bound method NDFrame.head of          Entity Code  Year  Average annual working ho
urs per worker
0     Argentina  ARG  1950                                2034.0000
1     Argentina  ARG  1951                                2037.8667
2     Argentina  ARG  1952                                2041.7408
3     Argentina  ARG  1953                                2045.6223
4     Argentina  ARG  1954                                2049.5112
...         ...  ...   ...                                      ...
3465    Vietnam  VNM  2013                                2267.4883
3466    Vietnam  VNM  2014                                2150.6357
3467    Vietnam  VNM  2015                                2169.5916
3468    Vietnam  VNM  2016                                2169.5916
3469    Vietnam  VNM  2017                                2169.5916

[3470 rows x 4 columns]>
```

### STEP3: Checking whether there exist any null values in the data set

In [5]:
```
Working_hours_data.isnull()
```

Out[5]:

| | Entity | Code | Year | Average annual working hours per worker |
|---|---|---|---|---|
| **0** | False | False | False | False |
| **1** | False | False | False | False |
| **2** | False | False | False | False |
| **3** | False | False | False | False |
| **4** | False | False | False | False |
| **...** | ... | ... | ... | ... |
| **3465** | False | False | False | False |
| **3466** | False | False | False | False |
| **3467** | False | False | False | False |
| **3468** | False | False | False | False |
| **3469** | False | False | False | False |

3470 rows × 4 columns

There is no missing values in the data set

## STEP4: Checking informations regarding our data set like data types and dimensions

In [6]:
```
Working_hours_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3470 entries, 0 to 3469
Data columns (total 4 columns):
 #   Column                                  Non-Null Count  Dtype  
---  ------                                  --------------  -----  
 0   Entity                                  3470 non-null   object 
 1   Code                                    3470 non-null   object 
 2   Year                                    3470 non-null   int64  
 3   Average annual working hours per worker  3470 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 108.6+ KB
```

We are having 3470 entries across 4 columns . Entity and Code having data type object , Year has integer and Average annnual working hours per worker in float

## STEP5: We will describe our data set with basic statistics

In [7]:
```
Working_hours_data.describe()
```

Out[7]:

| | Year | Average annual working hours per worker |
|---|---|---|
| **count** | 3470.000000 | 3470.000000 |
| **mean** | 1986.706052 | 2005.061167 |
| **std** | 23.502962 | 303.706114 |
| **min** | 1870.000000 | 1353.886800 |
| **25%** | 1973.000000 | 1805.052375 |
| **50%** | 1991.000000 | 1976.041000 |
| **75%** | 2004.000000 | 2172.774525 |
| **max** | 2017.000000 | 3483.000000 |

From above we notice that mean working hours is 2005 annualy and maximum it has reached to 3483 .

# We will prepare four different bar charts for our data set

1. Comparing different country's mean working hours for all years.
2. Comparing differnt country's working hours for most recent year available.
3. Comparing within Country how the woking hours changed by different years.
4. Comparing top 5 countries for past four years annual working hours.

## 1. Comparing different Country's mean working hours for all years.

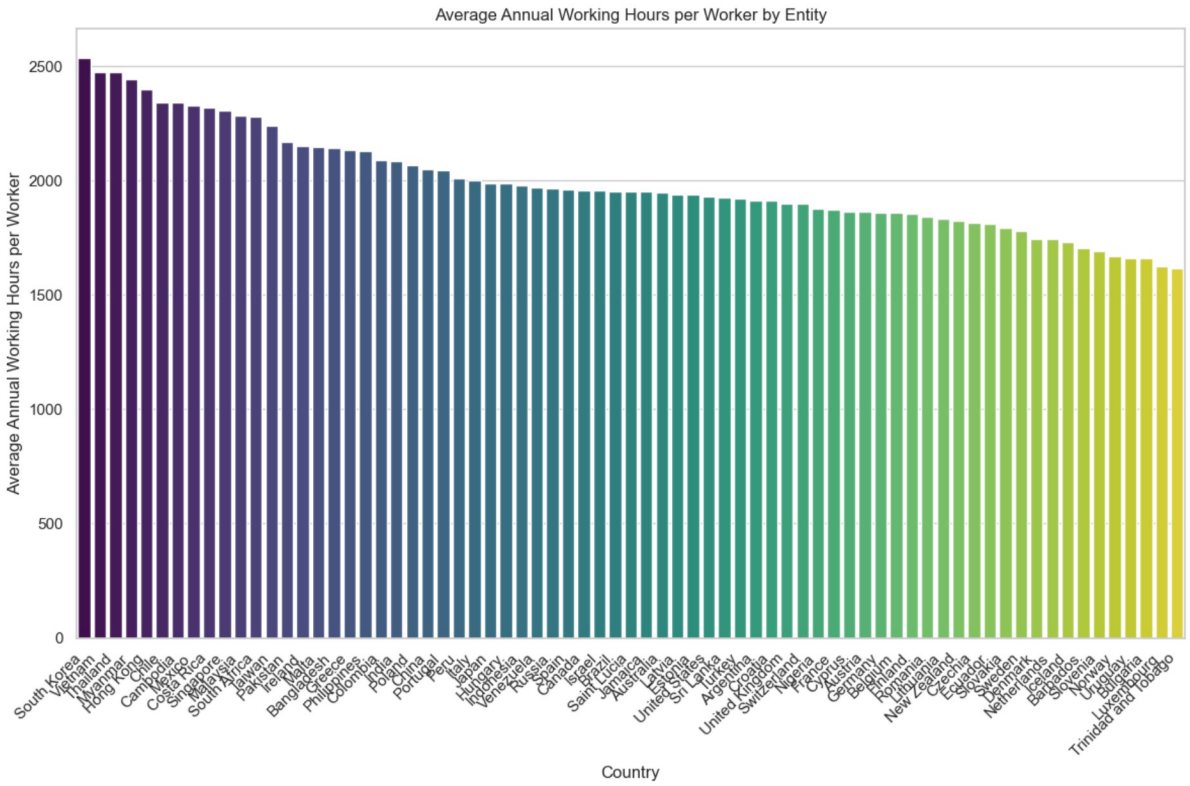
### STEP1: Group by Country for different year and taking mean working hours.

In [16]:
```
group_by_country = Working_hours_data.groupby(['Entity'])['Average annual working h
group_by_country_desc = group_by_country.sort_values(by='Average annual working hou
print(group_by_country_desc)
```

```
                Entity  Average annual working hours per worker
57          South Korea                                2535.875423
70              Vietnam                                2474.529052
63             Thailand                                2473.429988
40              Myanmar                                2445.061081
24            Hong Kong                                2398.650557
..                  ...                                      ...
44               Norway                                1670.351426
68              Uruguay                                1659.265950
7              Bulgaria                                1658.840770
36           Luxembourg                                1623.142075
64  Trinidad and Tobago                                1616.944917

[71 rows x 2 columns]
```

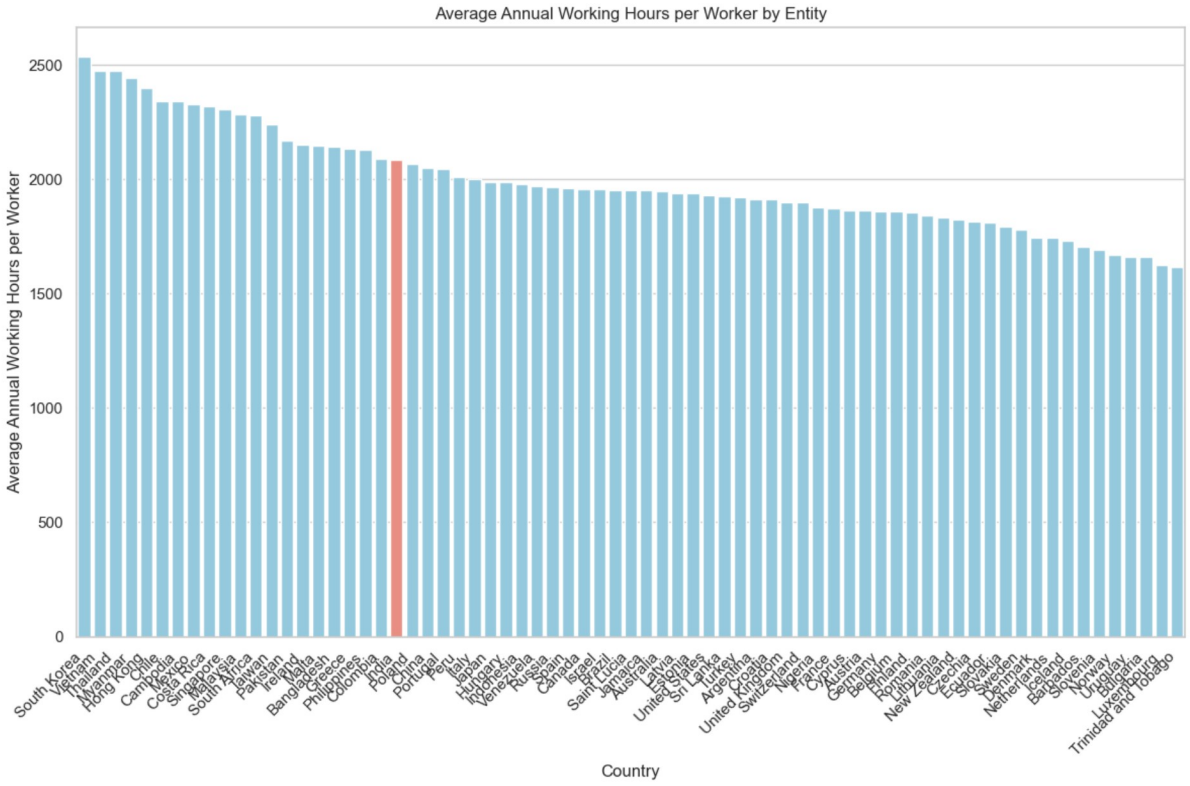
## STEP2:Visualising a appealing bar chart using libraries matplotlib and seaborn.

In [21]:
```
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")

plt.figure(figsize=(12, 8))
ax = sns.barplot(x='Entity', y='Average annual working hours per worker', data=grou
ax.set(xlabel='Country', ylabel='Average Annual Working Hours per Worker', title='A
plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```




## STEP3: Visualising a bar chart in which india position could be seen easily among all countries.

In [22]:
```
highlight_color = ['skyblue' if country != 'India' else 'salmon' for country in gro

sns.set(style="whitegrid")

plt.figure(figsize=(12, 8))
ax = sns.barplot(x='Entity', y='Average annual working hours per worker', data=grou
ax.set(xlabel='Country', ylabel='Average Annual Working Hours per Worker', title='A
plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```




CONCLUSION:

1. South korea has highest mean Annual working hours per worker.

2.Trinidad and lobargo has least mean Annual working hours per worker.

3.India ranks 21 for mean Annual working hours per worker.

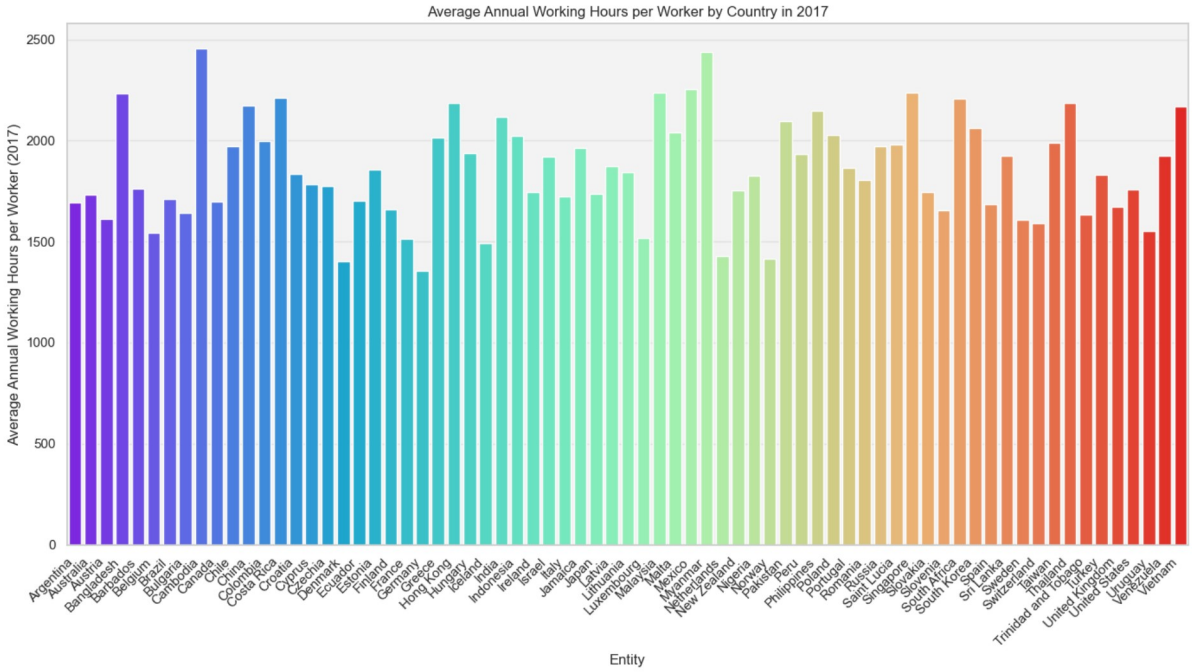
## 2. Comparing differnt country's working hours for most recent year available..

STEP1: Group by Country for most recent year available (2017) and taking working hours.

In [ ]:
```
country_2017 = Working_hours_data.groupby('Entity').last().reset_index()
print(country_2017[['Entity', 'Average annual working hours per worker']])
```

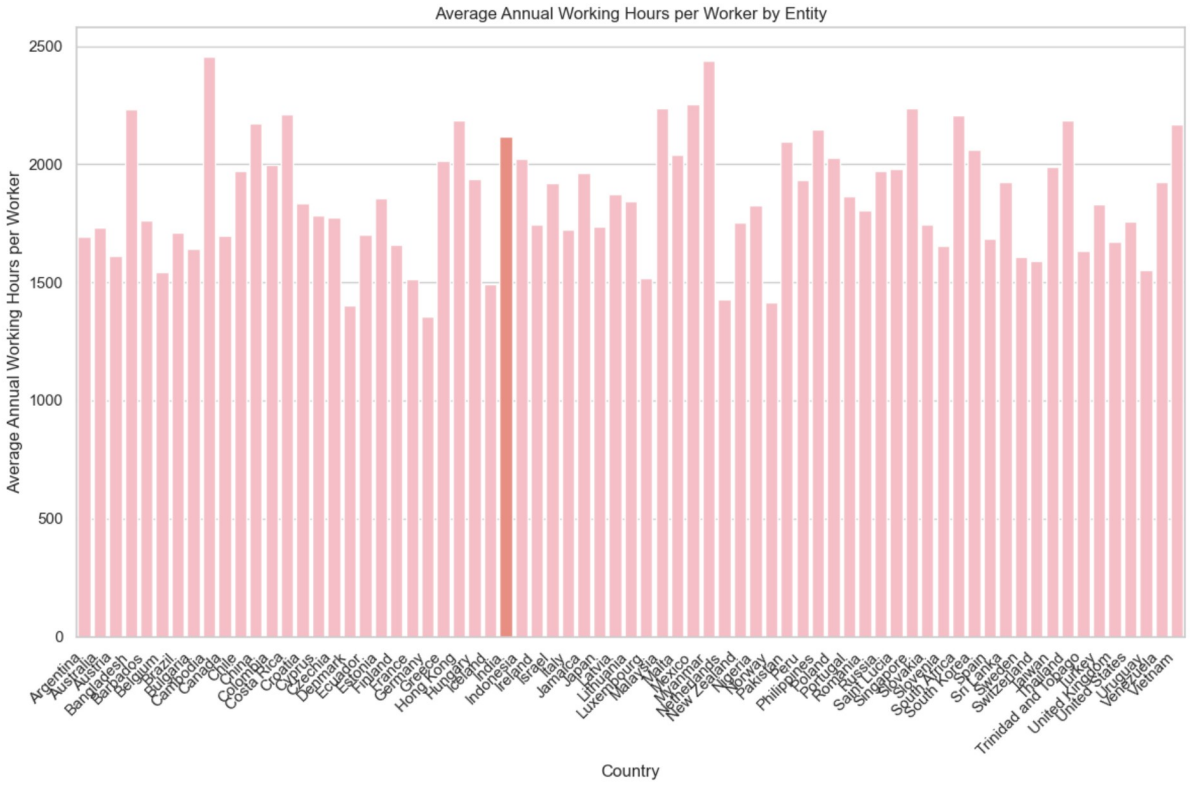
STEP2:Visualising the country's annual working hours for the most recent year 2017

In [33]:
```
sns.set(style="whitegrid", rc={"axes.facecolor": "#f3f3f3", "grid.color": "#e1e1e1"

plt.figure(figsize=(14, 8))
ax = sns.barplot(x='Entity', y='Average annual working hours per worker', data=coun
ax.set(xlabel='Entity', ylabel='Average Annual Working Hours per Worker (2017)', ti

ax.yaxis.grid(True)

plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```




## STEP3: Visualising a bar chart in which india position could be seen easily among all countries for year 2017.

In [41]:
```
highlight_color = ['lightpink' if country != 'India' else 'salmon' for country in c

sns.set(style="whitegrid")

plt.figure(figsize=(12, 8))
ax = sns.barplot(x='Entity', y='Average annual working hours per worker', data=coun
ax.set(xlabel='Country', ylabel='Average Annual Working Hours per Worker', title='A
plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```




CONCLUSION

1.For year 2017 Argentina has highest annual working hours per worker.

2.For year 2017 Vietnam has lowest annual working hours per worker.

3.For year 2017 India ranks 28 for annual working hours per worker

## 3. Comparing within Country how the woking hours changed by different years.

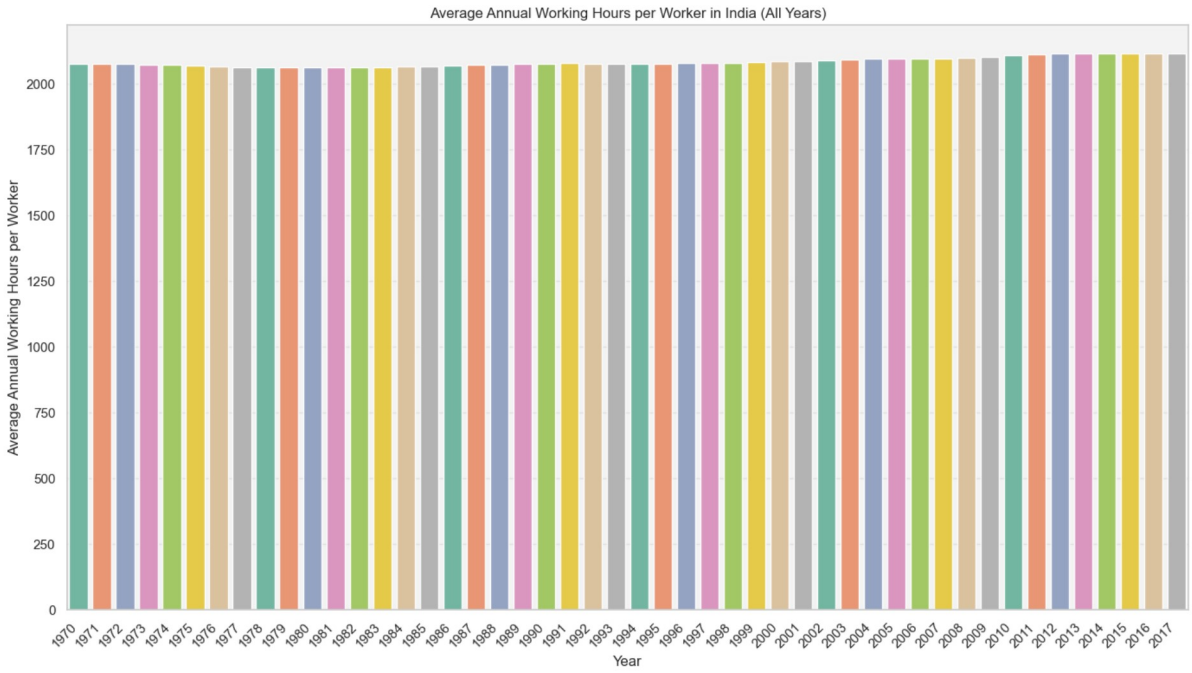
STEP1:Comparing annual working hours for india through the years .

In [45]:
```
India_data = Working_hours_data[Working_hours_data['Entity'] == 'India']
print(India_data)
```

| | Entity | Code | Year | Average annual working hours per worker |
|---|---|---|---|---|
| 1377 | India | IND | 1970 | 2077.3308 |
| 1378 | India | IND | 1971 | 2077.0852 |
| 1379 | India | IND | 1972 | 2076.5603 |
| 1380 | India | IND | 1973 | 2074.8291 |
| 1381 | India | IND | 1974 | 2072.3657 |
| 1382 | India | IND | 1975 | 2069.9565 |
| 1383 | India | IND | 1976 | 2067.1599 |
| 1384 | India | IND | 1977 | 2065.3796 |
| 1385 | India | IND | 1978 | 2064.4177 |
| 1386 | India | IND | 1979 | 2064.2246 |
| 1387 | India | IND | 1980 | 2064.0278 |
| 1388 | India | IND | 1981 | 2063.6230 |
| 1389 | India | IND | 1982 | 2064.1531 |
| 1390 | India | IND | 1983 | 2065.1204 |
| 1391 | India | IND | 1984 | 2066.7083 |
| 1392 | India | IND | 1985 | 2068.6077 |
| 1393 | India | IND | 1986 | 2070.5779 |
| 1394 | India | IND | 1987 | 2072.2705 |
| 1395 | India | IND | 1988 | 2075.0774 |
| 1396 | India | IND | 1989 | 2078.1321 |
| 1397 | India | IND | 1990 | 2078.4795 |
| 1398 | India | IND | 1991 | 2079.3042 |
| 1399 | India | IND | 1992 | 2076.9104 |
| 1400 | India | IND | 1993 | 2075.8694 |
| 1401 | India | IND | 1994 | 2075.6873 |
| 1402 | India | IND | 1995 | 2077.0972 |
| 1403 | India | IND | 1996 | 2079.8303 |
| 1404 | India | IND | 1997 | 2079.3826 |
| 1405 | India | IND | 1998 | 2080.7412 |
| 1406 | India | IND | 1999 | 2082.6938 |
| 1407 | India | IND | 2000 | 2086.6284 |
| 1408 | India | IND | 2001 | 2086.7727 |
| 1409 | India | IND | 2002 | 2088.8936 |
| 1410 | India | IND | 2003 | 2091.9958 |
| 1411 | India | IND | 2004 | 2095.7783 |
| 1412 | India | IND | 2005 | 2096.9160 |
| 1413 | India | IND | 2006 | 2097.0964 |
| 1414 | India | IND | 2007 | 2098.1213 |
| 1415 | India | IND | 2008 | 2099.6331 |
| 1416 | India | IND | 2009 | 2103.5361 |
| 1417 | India | IND | 2010 | 2110.3926 |
| 1418 | India | IND | 2011 | 2113.6245 |
| 1419 | India | IND | 2012 | 2115.7102 |
| 1420 | India | IND | 2013 | 2117.1030 |
| 1421 | India | IND | 2014 | 2118.0881 |
| 1422 | India | IND | 2015 | 2117.0117 |
| 1423 | India | IND | 2016 | 2117.0117 |
| 1424 | India | IND | 2017 | 2117.0117 |

## STEP2: Visualising Annual working hours for India past years .

In [65]:
```
sns.set(style="whitegrid", rc={"axes.facecolor": "#f3f3f3", "grid.color": "#e1e1e1"

plt.figure(figsize=(14, 8))
ax = sns.barplot(x='Year', y='Average annual working hours per worker', data=India_

ax.set(xlabel='Year', ylabel='Average Annual Working Hours per Worker', title='Aver

ax.yaxis.grid(True)

plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```




CONCLUSION

1.IN year 1970 annual working hours per worker was 2077.

2.In year 2017 annual working hours per worker was 2117.

3.We can easily observe that in India annual working hours per worker increased from 1970 to 2017.

## 4. Comparing top 5 countries for past four years annual working hours.

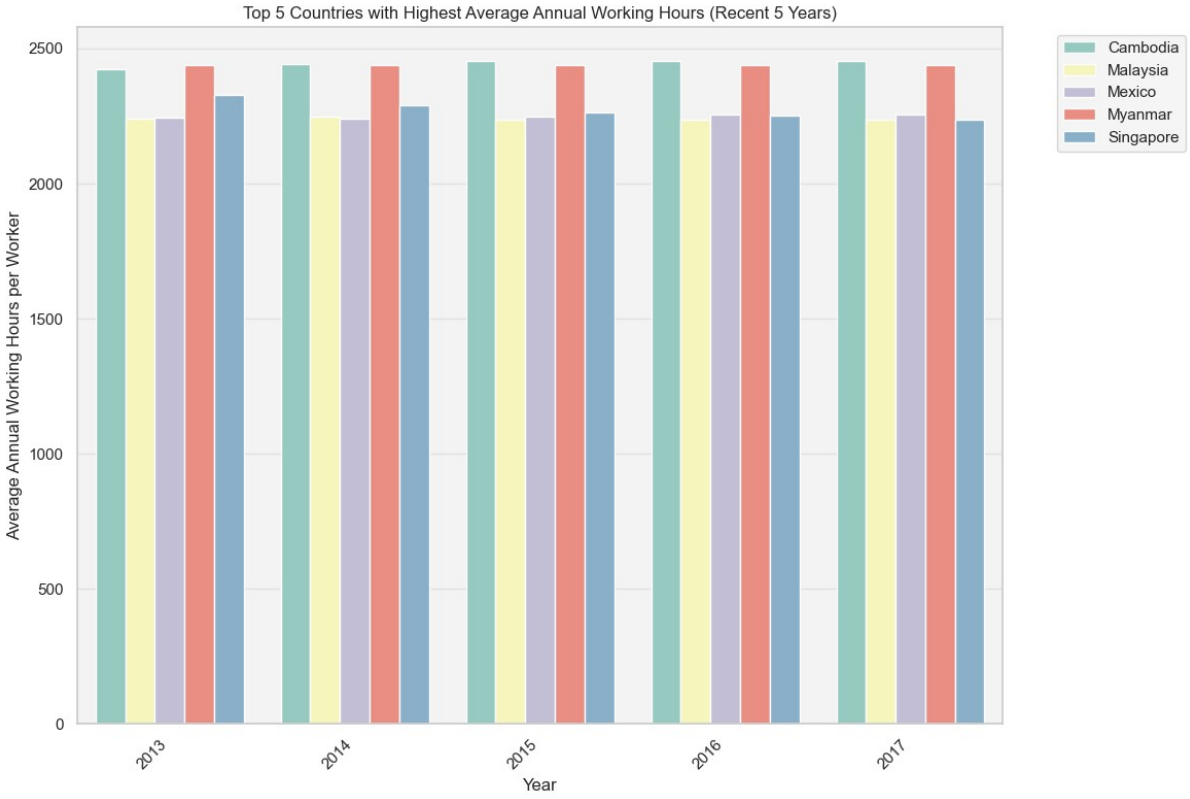
STEP1: Get the top 5 countries with the highest average annual working hours for past four years.

In [66]:
```
recent_years_data = Working_hours_data[ Working_hours_data['Year'] >=  Working_hour

top_countries = recent_years_data.groupby('Entity')['Average annual working hours p

top_countries_data = recent_years_data[recent_years_data['Entity'].isin(top_countri
print(top_countries_data)
```

| | Entity | Code | Year | Average annual working hours per worker |
|---|---|---|---|---|
| 456 | Cambodia | KHM | 2013 | 2423.9707 |
| 457 | Cambodia | KHM | 2014 | 2442.6956 |
| 458 | Cambodia | KHM | 2015 | 2455.5508 |
| 459 | Cambodia | KHM | 2016 | 2455.5508 |
| 460 | Cambodia | KHM | 2017 | 2455.5508 |
| 1882 | Malaysia | MYS | 2013 | 2240.9502 |
| 1883 | Malaysia | MYS | 2014 | 2249.6169 |
| 1884 | Malaysia | MYS | 2015 | 2238.2729 |
| 1885 | Malaysia | MYS | 2016 | 2238.2729 |
| 1886 | Malaysia | MYS | 2017 | 2238.2729 |
| 1974 | Mexico | MEX | 2013 | 2244.0000 |
| 1975 | Mexico | MEX | 2014 | 2242.0000 |
| 1976 | Mexico | MEX | 2015 | 2248.0000 |
| 1977 | Mexico | MEX | 2016 | 2255.0000 |
| 1978 | Mexico | MEX | 2017 | 2255.0000 |
| 2022 | Myanmar | MMR | 2013 | 2438.6584 |
| 2023 | Myanmar | MMR | 2014 | 2438.1577 |
| 2024 | Myanmar | MMR | 2015 | 2437.8633 |
| 2025 | Myanmar | MMR | 2016 | 2437.8633 |
| 2026 | Myanmar | MMR | 2017 | 2437.8633 |
| 2602 | Singapore | SGP | 2013 | 2328.0503 |
| 2603 | Singapore | SGP | 2014 | 2290.3225 |
| 2604 | Singapore | SGP | 2015 | 2263.1465 |
| 2605 | Singapore | SGP | 2016 | 2252.5857 |
| 2606 | Singapore | SGP | 2017 | 2237.7263 |

## STEP2: Visualising the top 5 countries for past 4 years years wise.

In [71]:
```
sns.set(style="whitegrid", rc={"axes.facecolor": "#f3f3f3", "grid.color": "#e1e1e1"

plt.figure(figsize=(14, 8))
ax = sns.barplot(x='Year', y='Average annual working hours per worker', hue='Entity

ax.set(xlabel='Year', ylabel='Average Annual Working Hours per Worker', title='Top

ax.yaxis.grid(True)

plt.xticks(rotation=45, ha='right')

plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')


plt.tight_layout(rect=[0, 0, 0.85, 1])
plt.show()
```




CONCLUSION:

1.Top five leading countries for past five years in Annual working hours are Cambodia Malesia Mexico Myanmar Singapore

Above we analysed different Countries for Anuual working Hours Per Worker