QUEENSLAND ADVANCE

SCENARIO: The allocation of public money (obtained from the public via taxes) is a politically sensitive activity with governments regularly coming under scrutiny for how this money is spent. A respected media organisation is looking into the Queensland Government's Advance Queensland program. The resulting story could be a "good news" story reporting on the success of the program, however if inappropriate spending or irregularities are found, it could become a story that is critical of the scheme, and potentially the Government.

As a data analyst, your task is to analyse the publicly available data on the distribution of the funds over time. You are looking for patterns that may support the "good news" story, or which may be a cause for concern. It is up to you how deeply you explore the data, but at a minimum you should look at (a) the balance between South-East Queensland and the remainder of the state (regional Queensland); and (b) how distributions align with the objectives of the scheme which may include supporting specified groups of people.

ETHICAL APPROACH: You are expected to be fair and ethical in your analysis, and therefore the insights that you draw should take into account contextual factors. You should avoid simplistic assumptions like assuming all groups and activities should receive equal funding. For example, disproportionate funding may be appropriate due to social circumstances or the costs involved in a particular activity. Further, benefits to Queensland may come in different forms. For example, cultural benefits cannot be directly compared to economic benefits.

ESSENTIAL REQUIREMENTS: Your task as a data analyst is to:

- Ensure that you use the techniques and libraries/packages that have been used in class
- Identify high quality questions that when answered may be helpful in addressing the scenario above
- Obtain the data in JSON form from the API.
- Clean and filter the data as appropriate
- Analyse the data in a way that answers your questions and ultimately addresses the concern in the scenario
- Visualise your results in a meaningful way that is helpful in making visible key findings
- Provide a detailed summary of the insights found and how they address the original questions and scenario

AUTHENTICITY AND INTEGRITY: You will be marked on (a) *HOW* you undertake the task todgether; with (b) detail of *WHY* you made various decisions involved in the tasks; and (c) acknowledgement of **WHERE** you used material that is not directly yours. Therefore, you must document your thinking and approach throughout the notebook using the Markdown cells, and give credit to other resources as appropriate. You are encouraged to use the `Exemplars PDF` to help write your code. You may use online resources including `GenAI tools` and `stackoverflow` to help you write your code, however you must acknowledge that you are using these resources in the markdown cells explaining your analysis. Note that you do not need to use formal referencing for this.

```python
In [15]: import pandas as pd
         import plotly.express as px
         URL = "https://www.data.qld.gov.au/dataset/db190f2d-f866-4811-9a6e-4b78744b551b/resource/0f97b985-f5c7-49d2-8
         df = pd.read_csv(URL, encoding='latin-1')
         df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 16 columns):
 #   Column                                            Non-Null Count  Dtype
---  ------                                            --------------  -----
 0   Program                                           1339 non-null   object
 1   Round                                             1288 non-null   object
 2   Recipient Name                                    1339 non-null   object
 3   Physical Address of Recipient – Suburb/Location   1339 non-null   object
 4   Physical Address of Recipient – Post Code         1268 non-null   float64
 5   University Collaborator (if applicable)           12 non-null     object
 6   Other Partners; Collaborators (if applicable)     477 non-null    object
 7   Investment/Project Title                          1339 non-null   object
 8   Primary Location of Activity/Project – Suburb     1306 non-null   object
 9   Primary Location of Activity/Project – Post Code  1284 non-null   float64
 10  Multiple Locations of Activity/Project (if applicable)  249 non-null  object
 11  Approval date                                     1339 non-null   object
 12  Local Government /Council                         1339 non-null   object
 13  RAP Region                                        1339 non-null   object
 14  State Electorate                                  1339 non-null   object
 15  Actual Contractual Commitment ($)                 1339 non-null   object
dtypes: float64(2), object(14)
memory usage: 167.5+ KB
```

In [16]:
```python
ean Data:
ange data format
'Actual Contractual Commitment ($)']= df['Actual Contractual Commitment ($)'].replace(',', '',regex=True).ast
'Approval date'] =  pd.to_datetime(df["Approval date"], format= "%d/%m/%Y")

ill missing values in the specified columns with 'None'

umns_to_fill = ['Round ',
                'Physical Address of Recipient – Post Code',
                'University Collaborator (if applicable)',
                'Other Partners; Collaborators (if applicable)',
                'Primary Location of Activity/Project – Suburb',
                'Primary Location of Activity/Project – Post Code',
                'Multiple Locations of Activity/Project (if applicable)']

columns_to_fill] = df[columns_to_fill].fillna('None')
int(df.isna().sum())

dd YEAR and Month columns to future analysis:
['Year'] = df['Approval date'].dt.year
'Month'] = df['Approval date'].dt.month

dd another columns name south east and other region:
= ['Brisbane (C)', 'Ipswich (C)', 'Lockyer Valley (R)', 'Logan (C)', 'Moreton Bay (R)', 'Noosa (S)', 'Redlan
woom_SEQ = ['Toowoomba (R)']
 df['RAP Region'] == 'Gold Coast'

reate a boolean value to check if each value in 'Local Government /Council' is in SEQ
SEQ = df['Local Government /Council'].isin(SEQ)
SEQ_Toow = x.isin(Toowoom_SEQ)
'SEQ And Others'] = 'other areas'

se the boolean value to assign values to the 'SEQ And Others' column
loc[is_SEQ, 'SEQ And Others'] = 'SEQ'
loc[is_SEQ_Toow, 'SEQ And Others'] = 'SEQ'

hange all the recipient who named QUT or UQ into Queensland University of Technology and University of Queens
'Recipient Name '] = df.loc[df['Recipient Name'].str.contains('Queensland University of Technology'), 'Recipi
'Recipient Name ']= df.loc[df['Recipient Name'].str.contains('University of Queensland'), 'Recipient Name']=

heck other errors and turn it into empty
tern = r'^\s*$'
 row in df:
 df[row] = df[row].replace(pattern, '_empty_', regex = True)
```

In [17]:
```python
# The whole state sistuation:
state_by_year = df.groupby('Year')['Actual Contractual Commitment ($)'].sum()
px.line(state_by_year, y = 'Actual Contractual Commitment ($)',
        title = 'FUNDING OF QUEENSLAND ADVANCE BY YEARs',
        labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)"},
        text = 'Actual Contractual Commitment ($)',
       template="simple_white")
```

FUNDING OF QUEENSLAND ADVANCE BY YEARs
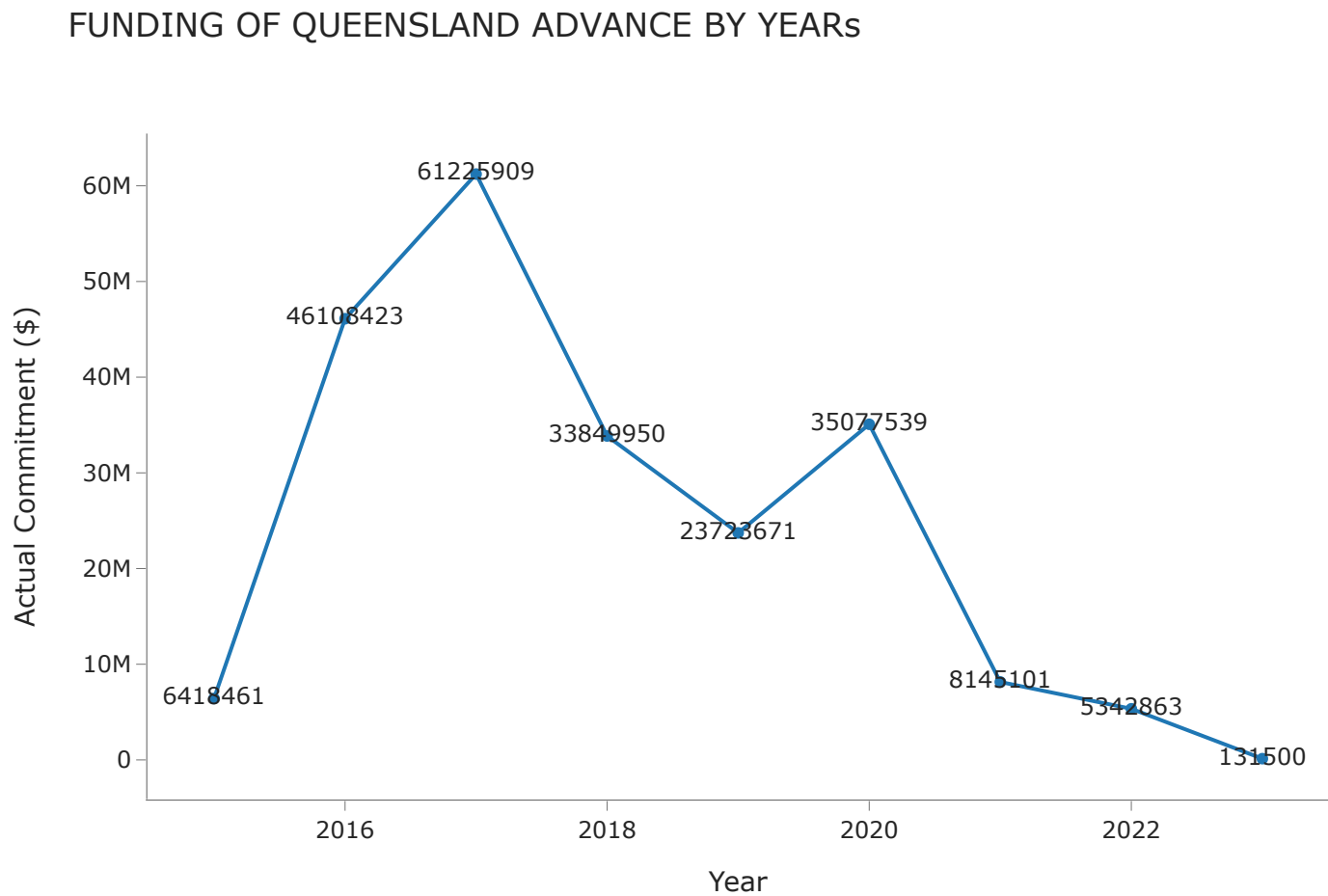
Introduction:

- The funding has been decreasing since 2018, but the decline became more serious in 2021 and 2022 due to the impact of the Covid situation. Given the consistent lack of funds over the years, this situation combine with the down trend of economic in the wordwide require goverment must be using their money more efficient than the past. That's why this report has been conducted to see if their fund being used effective or not.
- This report will examine three key aspects. Firstly, it will analyse the distribution of funds between Southeast Queensland and other regions. Secondly, it will focus on identifying special groups receiving support each year by examining which funds are being allocated in a given year. Finally, it will delve into the situation of Queensland over the past two years (2022 - 2023).

I. QUEENSLAND ADVANCE'S FUNDING SITUATION

## a. AMOUNT OF FUND BY REGIONS

```
In [18]:   funding by years:

l Contractual Commitment ($)',

And Others',
e_map={"SEQ": "lightskyblue", "other areas": "coral"},
rue,

rs={"Years": ["2015", "2016", "2017", "2018", "2019", "2020", "2021", "2022", "2023"], "Region": ["SEQ", "Oth
ple_white",
tual Contractual Commitment ($)": "Actual Commitment ($)"},
 DISTRIBUTION BETWEEN SOUTH EAST QUEENSLAND AND OTHER AREAS'
```

FUND DISTRIBUTION BETWEEN SOUTH EAST QUEENSLAND AND OTHER AREAS
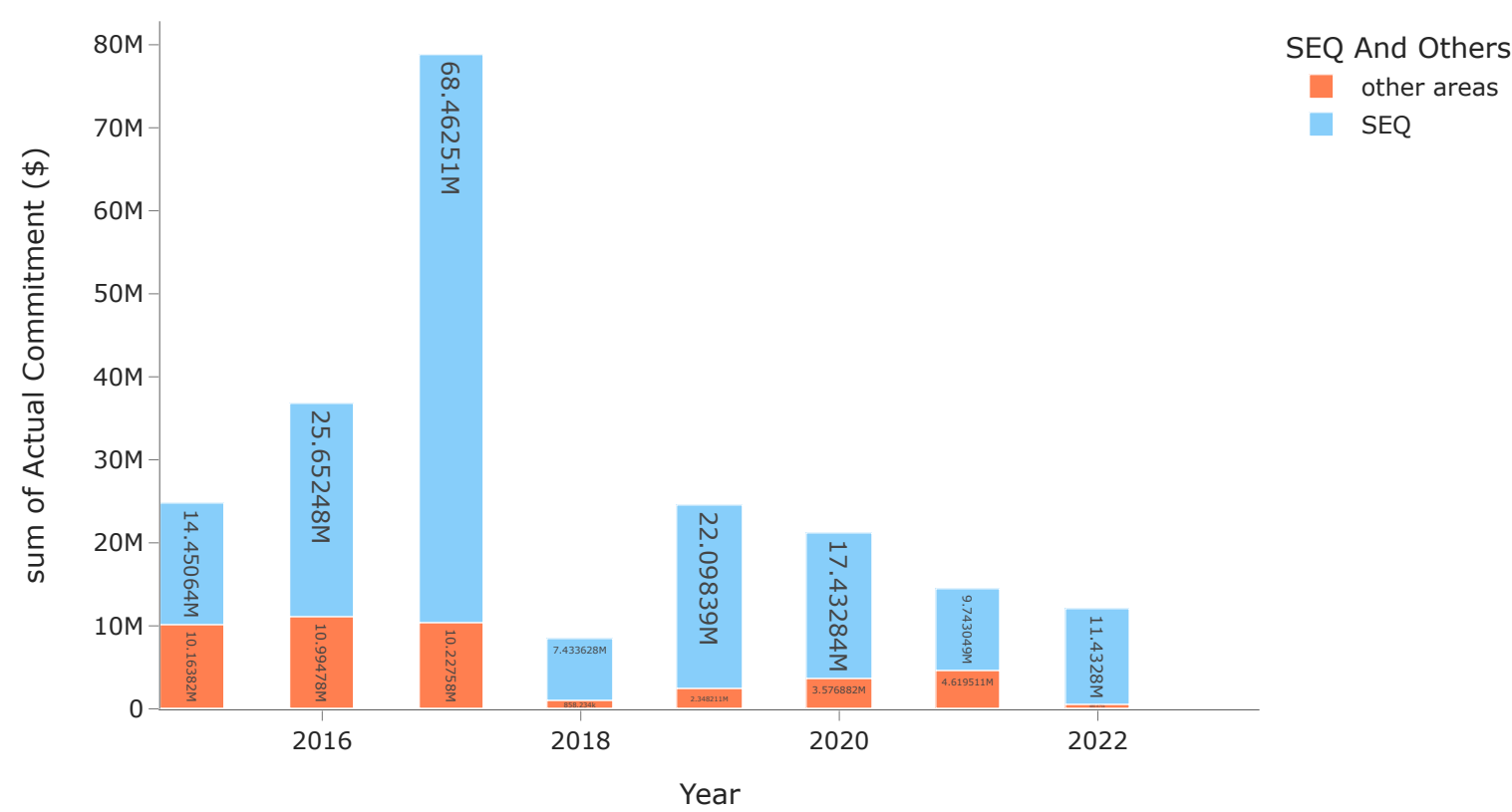
In [19]: *Group by regions:*
```
EQ_others = df.groupby('SEQ And Others')['Actual Contractual Commitment ($)'].sum()
ig = px.bar(SEQ_others, y='Actual Contractual Commitment ($)',
            title='FUND COMMITED BY LOCATION (SEQ AND OTHERS)', text_auto = True,
             template="simple_white",
            labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)", 'SEQ And Others':'Region'}
            )
ig.show()
```

FUND COMMITED BY LOCATION (SEQ AND OTHERS)
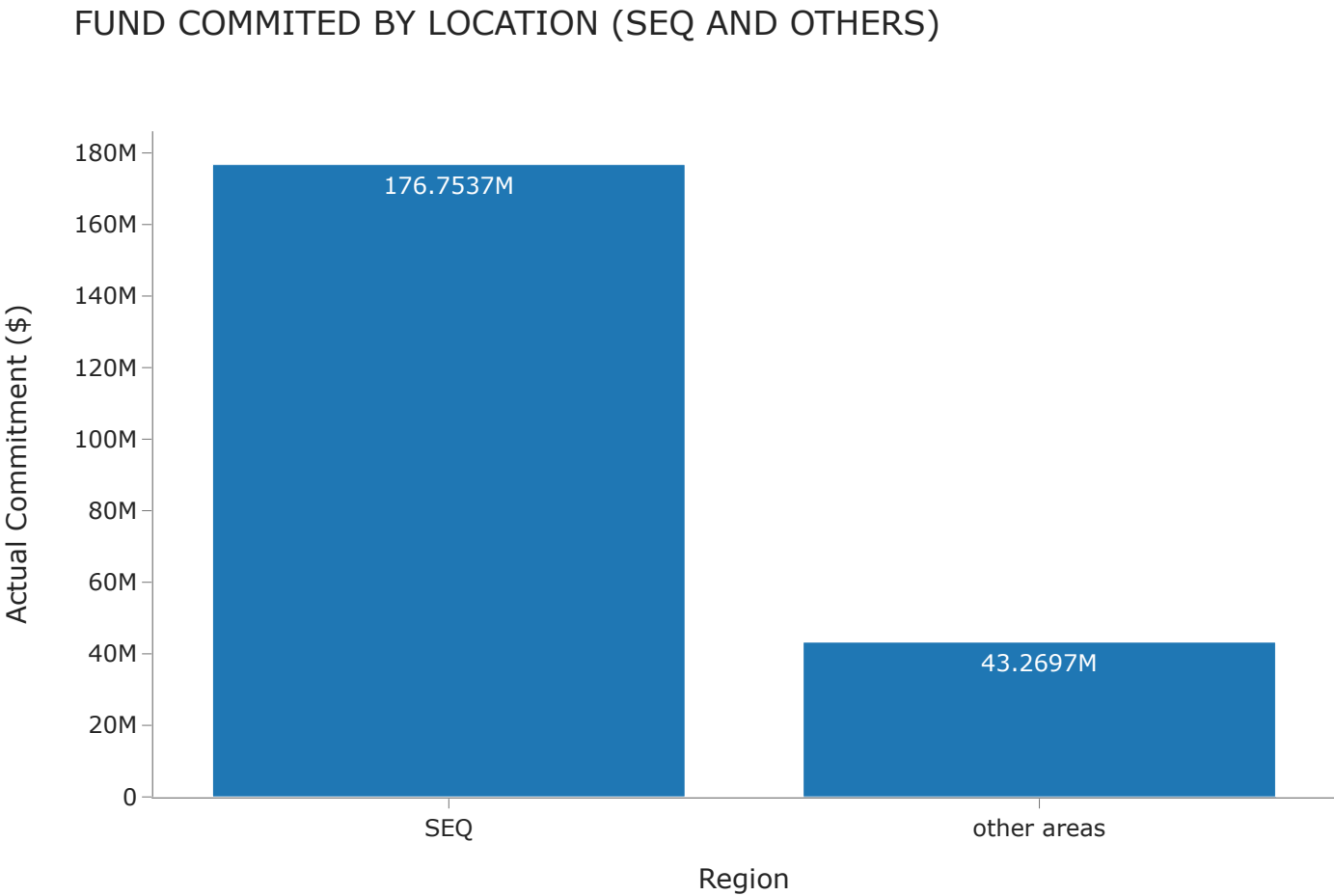
Insights:

- Southeast Queensland (SEQ) including Brisbane (C), Ipswich (C), Lockyer Valley (R), Logan (C), Moreton Bay (R), Noosa (S), Redland (C), Scenic Rim (R), Sunshine Coast (R), Gold Coast (C), Somerset (R).
- The fund spending for Southeast Queensland always more than another regions almost four times.
- The reason is SEQ is the area has the most poplation and also consider as head of Queensland. That is why SEQ get more funds than another regions

b. TOP FIVE SUBURDS GOT THE MOST INVESTMENT

In [20]:
```python
# Select the top five suburbs with the most investment
top_five_suburb = df.groupby('Physical Address of Recipient – Suburb/Location')['Actual Contractual Commitmen

# Reset the index to convert the Series to a DataFrame
top_five_suburbs_df = top_five_suburb.reset_index()

# Create a bar chart
fig = px.bar(top_five_suburbs_df,
             x='Physical Address of Recipient – Suburb/Location',
             y='Actual Contractual Commitment ($)',
             title='TOP FIVE SUBURBS WITH THE MOST INVERSTMENT ',
             labels={'Physical Address of Recipient – Suburb/Location': 'Suburb/Location', 'Actual Contractua
             text_auto = True, template="simple_white")

# Show the plot
fig.show()
# find the resone why this surburds got invest?
```

TOP FIVE SUBURBS WITH THE MOST INVERSTMENT
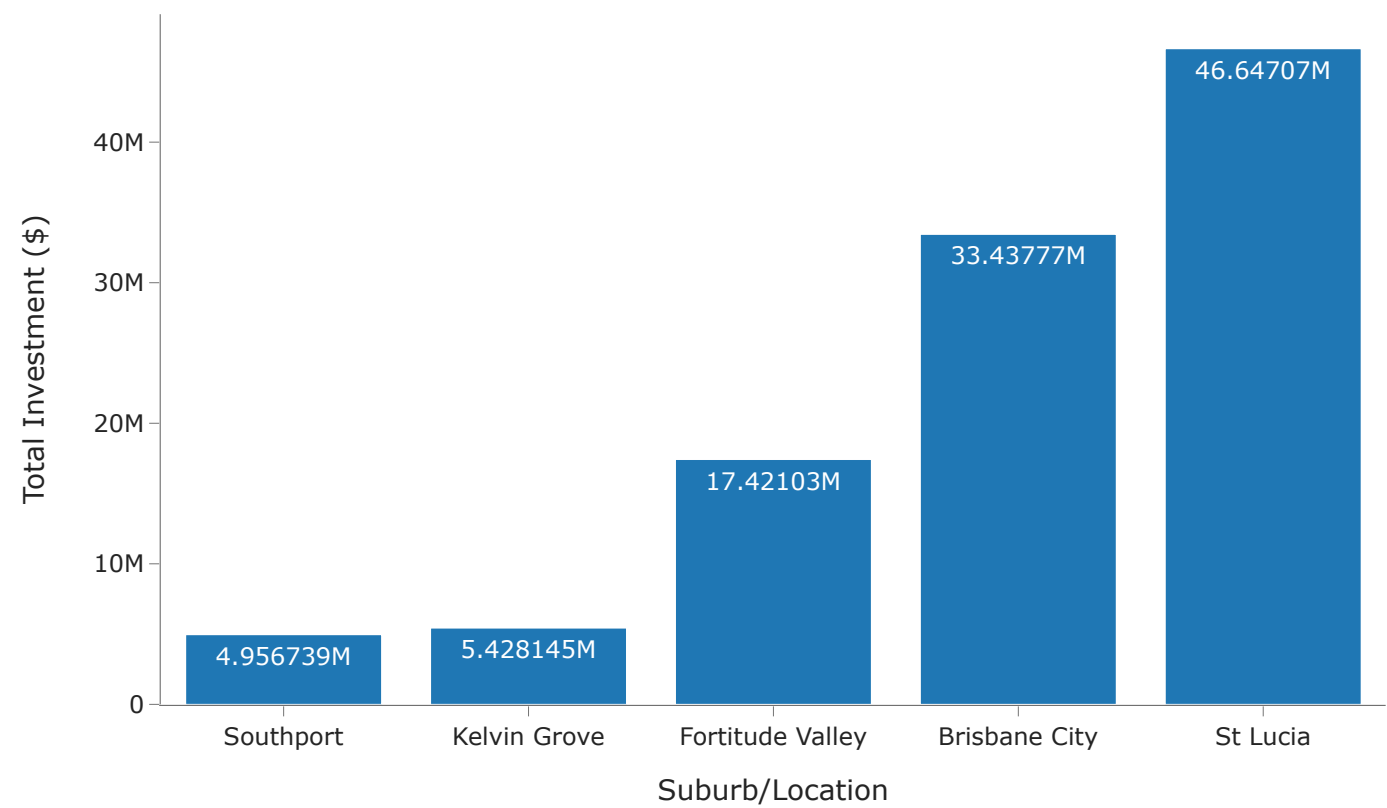
Insight:

- St Lucia, Brisbane City and Kelvin Grove are three suburbs that have big universities in Queensland.
- They receive more funds, which appear to be funding strategies that spend more on research and education of the government.
- Fortitude Valley are the places has a lot of start ups and company headquaters of Queensland.

II. FUNDING STRATEGIES:

a. BIGGEST FUND EVERY YEARS

In [21]:
```python
#Big Fund by year:
# Add YEAR and Month columns to future analysis:
df ['Year'] = df['Approval date'].dt.year
df['Month'] = df['Approval date'].dt.month
# Group the DataFrame by the combination of 'Recipient Name' and 'Year', and sum up the 'Actual Contractual C
grouped_df = df.groupby(['Program', 'Year'])['Actual Contractual Commitment ($)'].sum()

# Reset the index to convert the groupby result into a DataFrame
grouped_df = grouped_df.reset_index()

# Find the row with the maximum sum for each year
max_funds_each_year = grouped_df.loc[grouped_df.groupby('Year')['Actual Contractual Commitment ($)'].idxmax()

# Print the result
px.bar(max_funds_each_year,y = 'Actual Contractual Commitment ($)',
       x = 'Year',
       color = 'Program',
       text_auto = True,
       title = "BIGGEST FUNDS BY YEAR",
       labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)"},
       text = 'Actual Contractual Commitment ($)',
     template="simple_white")
```
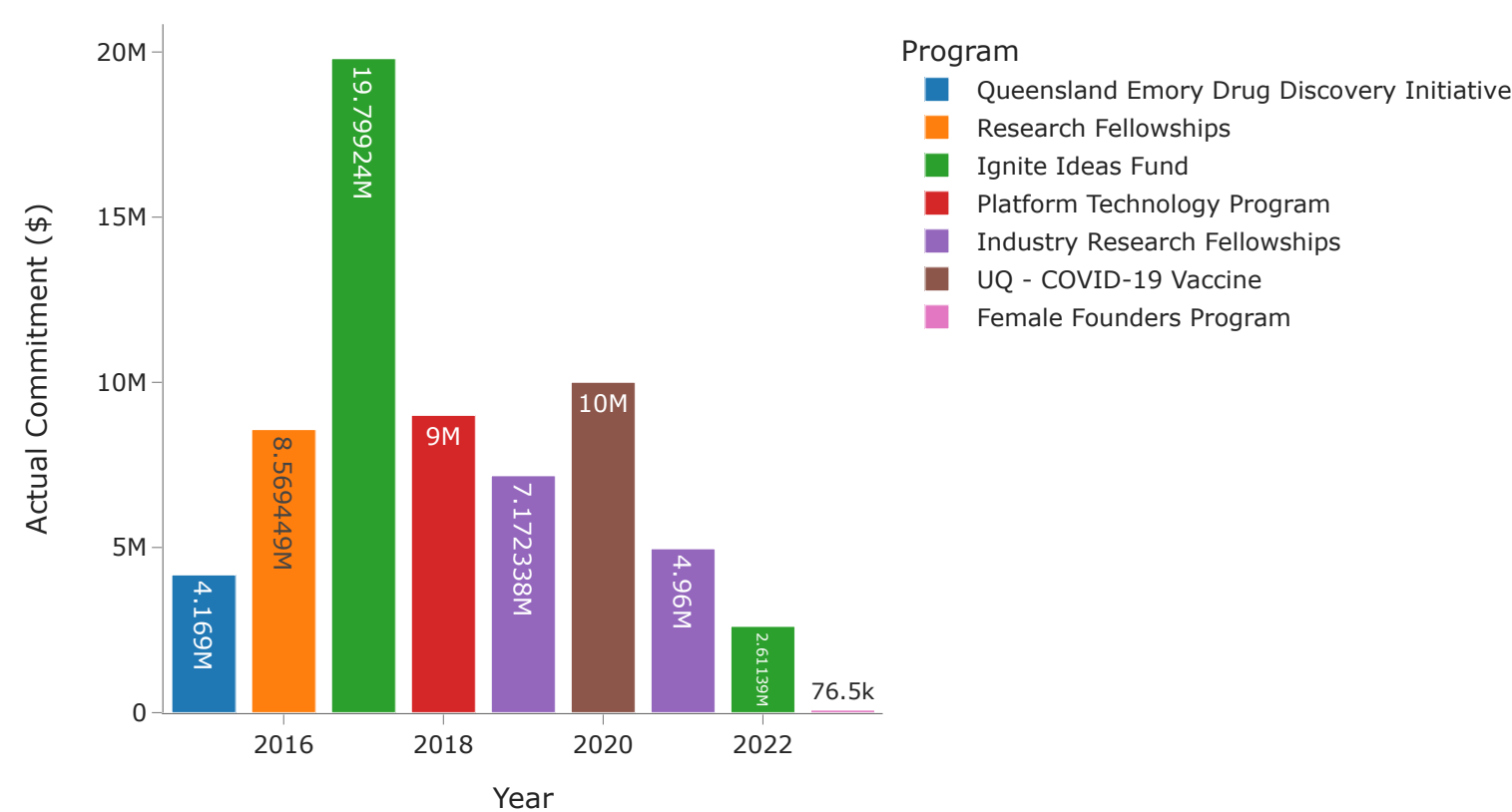
BIGGEST FUNDS BY YEAR

b. TOP HIGHEST - LOWEST FUNDS

In [22]:
```python
#five_highest:
five_high_programs = df.groupby('Program')['Actual Contractual Commitment ($)'].sum().sort_values().tail(5)
px.bar(five_high_programs,
        y = 'Actual Contractual Commitment ($)',
        text_auto = True,
        labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)", 'Program': 'Program name'},
        template="simple_white",
        title = 'FIVE HIGHEST SPENDING FUNDS')
```

FIVE HIGHEST SPENDING FUNDS
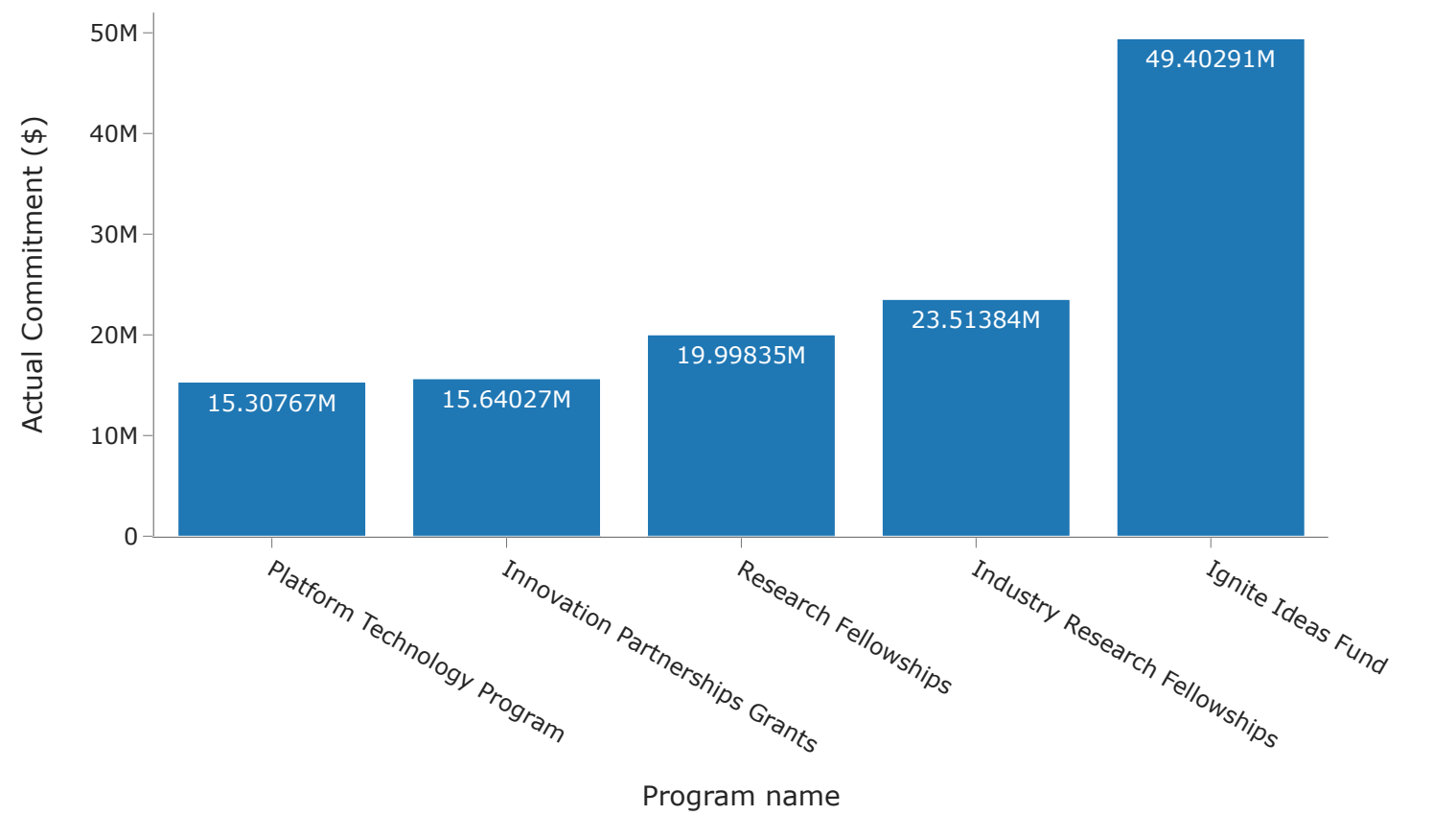
In [23]:
```python
#five_lowest:
five_lowest_programs = df.groupby('Program')['Actual Contractual Commitment ($)'].sum().sort_values().head(5)
px.bar(five_lowest_programs,
        y = 'Actual Contractual Commitment ($)',
        text_auto = True,
        labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)",
                'Program': 'Program name'},
        template="simple_white",
       title = 'FIVE LOWEST SPENDING FUNDS')
```

FIVE LOWEST SPENDING FUNDS
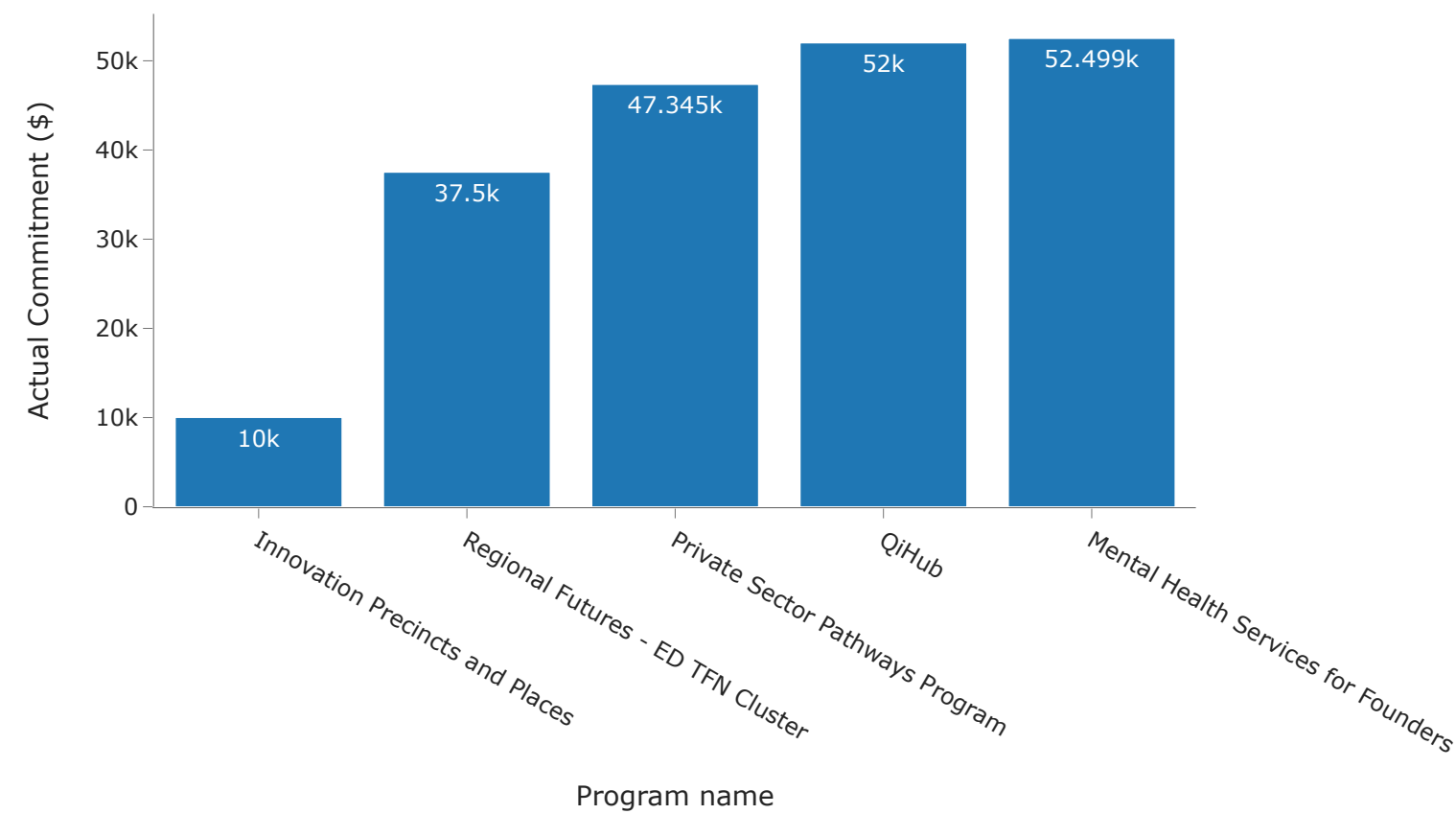
***Insights*:**

> The name of the fund represents the strategies invested of the Queensland goverment each year:
>
> - 2015: This fund aims to research and find a new medicine.
> - 2016: This fund is also a research fund
> - 2018: Support for technology projects makes economic growth
> - 2019: This fund is also a research fund
> - 2020: Fund for vaccinated COVID 19
> - 2021: This fund is also a research fund
> - 2022 & 2017: for small and medium business
> - Over the years, Queensland Advance's strategy is likely to spend funds on research and education.
> - The fund spending the most is the 'Ignite Ideas Fund'— a fund for small and medium businesses.
> - The limitation is that they do not consider regional and mental health fields to focus on.

c. TOP 5 RECIPIENTS:

```
In [24]: #CHECK IP ANY DUPLICATE VALUES IN THE RECIPIENT NAME MATCH WITH TOP 5
         df.loc[df['Recipient Name'].str.contains('Queensland University of Technology'), 'Recipient Name'] = 'Queensl
         df.loc[df['Recipient Name'].str.contains('University of Queensland'), 'Recipient Name'] = 'University of Quee
         df.loc[df['Recipient Name'].str.contains('Griffith University'), 'Recipient Name'] = 'Griffith University'
         df.loc[df['Recipient Name'].str.contains('Commonwealth Scientific And Industrial Research Organisation T/A Da
         df.loc[df['Recipient Name'].str.contains('Boeing Defence Australia Ltd'), 'Recipient Name'] = 'Boeing Defence
```

```
In [25]: # WHO GOT MOST MONEY :
         five_highest_repicients = df.groupby('Recipient Name')['Actual Contractual Commitment ($)'].sum().sort_values
         px.bar(five_highest_repicients,
                title= 'TOP 5 RECIPIENTS GOT COMMITMENT',
                text_auto = True,
                labels={'Actual Contractual Commitment ($)': 'Amount of Money ($)'},
                template="simple_white"
               )
```

TOP 5 RECIPIENTS GOT COMMITMENT
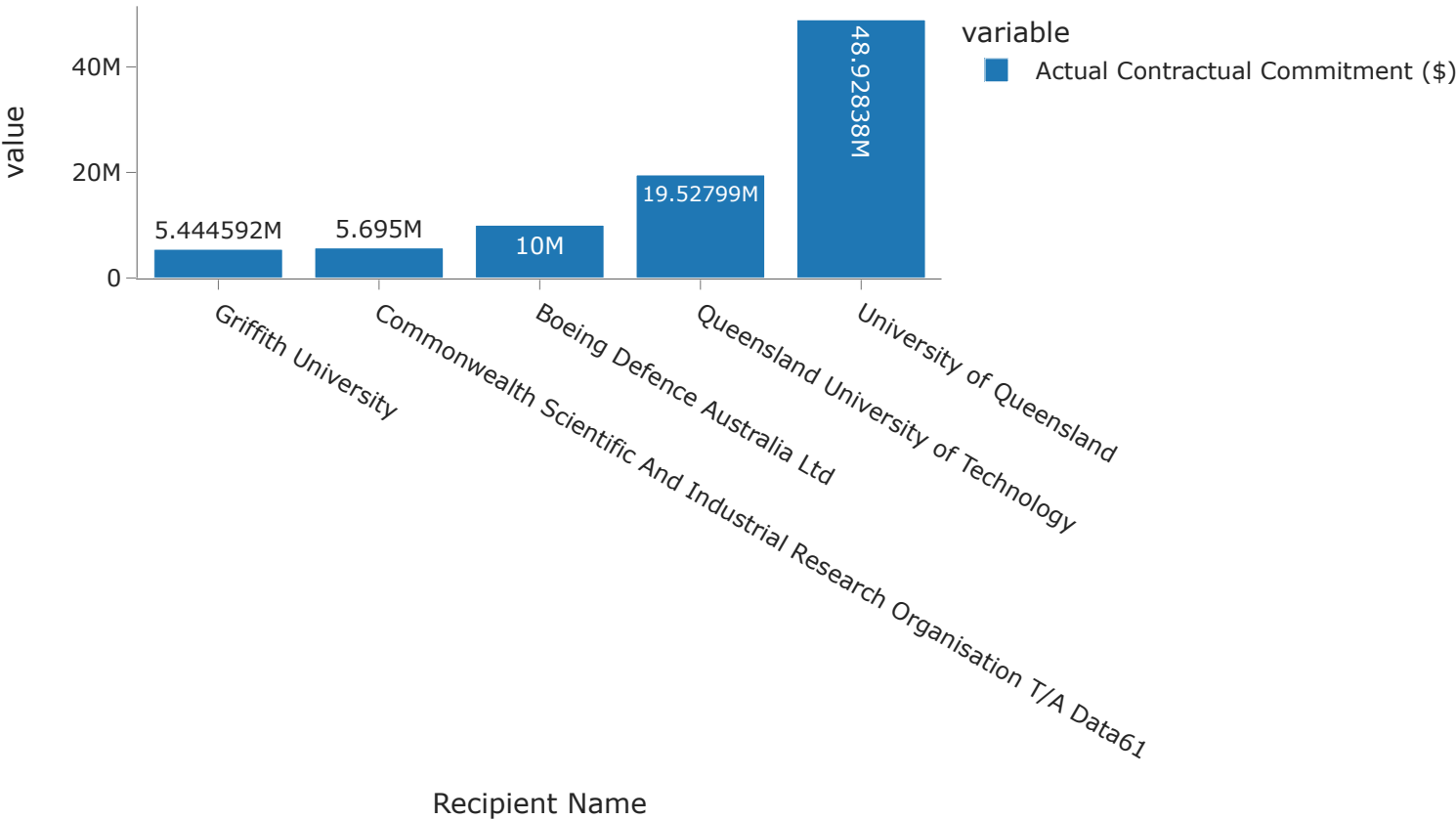
Insights

> - The University of Queensland, Queensland University of Technology and Griffith are three over 5 organisations being funding the most by Queensland Advance.
> - This insight strengthen the idea that Queensland's goverment investment strategy is more focus on research and education in the period 2015 to 2023.

III. FUNDING SITUATION IN 2023

In [26]:
```python
# choose data in the term 22-23 to analysis:
df_22_23 = df[(df['Approval date'].dt.year >= 2022) & (df['Approval date'].dt.month <= 3) & (df['Approval dat
df_22_23_table = pd.DataFrame(df_22_23.groupby(['SEQ And Others','Year', 'Month'])['Actual Contractual Commit
df_22_23_table
```

Out[26]:

SEQ And Others	Year	Month	Actual Contractual Commitment ($)
SEQ	2022	1	30000.0
		3	80000.0
	2023	1	30000.0
		3	76500.0
other areas	2022	2	35391.0
		3	51136.0
	2023	2	15000.0
		3	10000.0

In [27]:
```python
#situation in 2023 compare with same periods 2022:
px.histogram(df_22_23,
            x = 'Month',
            y = 'Actual Contractual Commitment ($)',
            color = 'Year',
            nbins = 6,
            text_auto = True,
            title = 'FUNDING BY YEAR BETWEEN Q1/2022 AND Q1/2023',
            template="simple_white",
            labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)"}
            )
```

FUNDING BY YEAR BETWEEN Q1/2022 AND Q1/2023
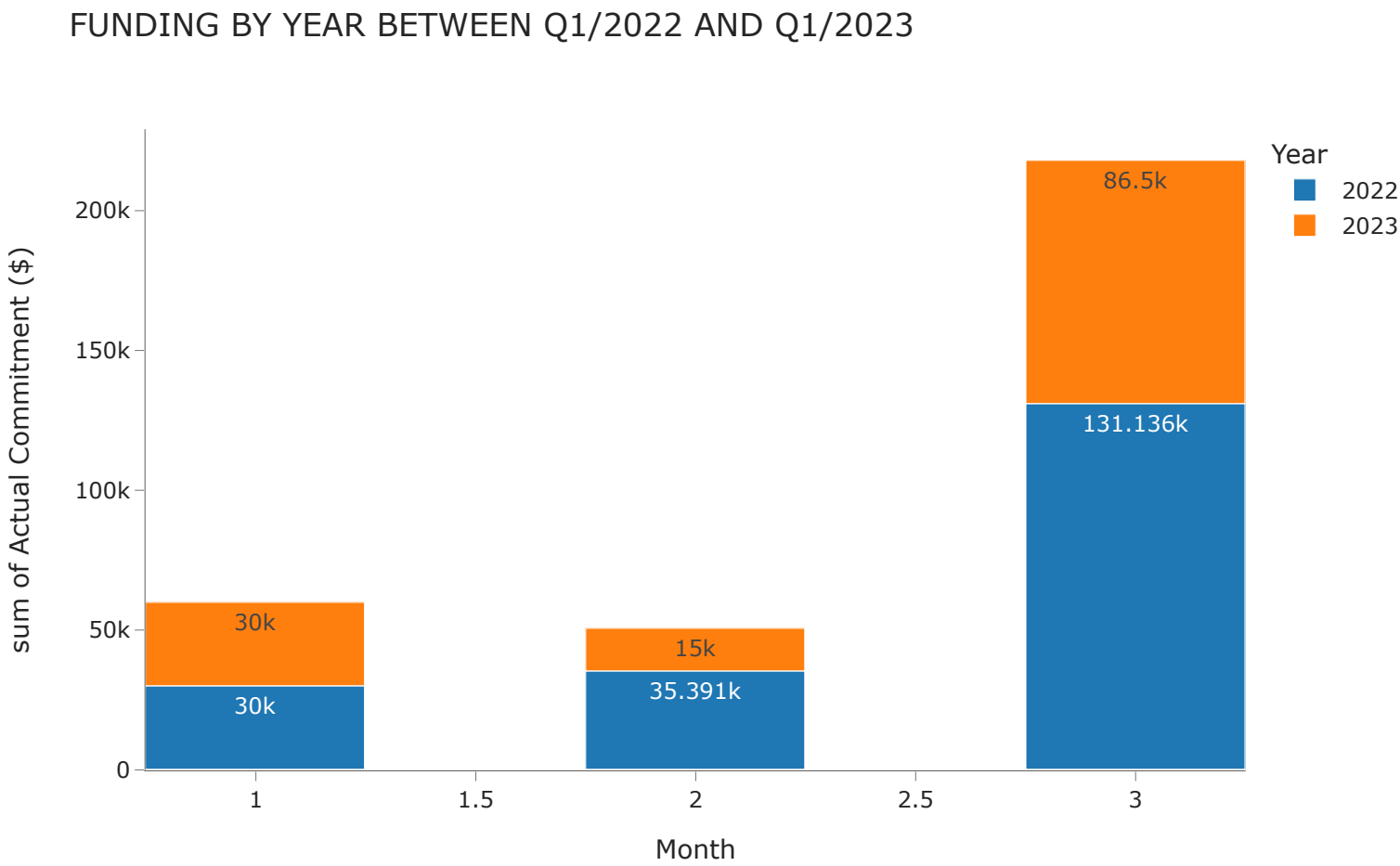
In [28]:
```python
px.bar(df_22_23,
        y = 'Actual Contractual Commitment ($)',
        x = 'Year',
        color = 'SEQ And Others',
        text_auto = True,
        title = 'FUND BY AREAS IN THE PERIOD Q1/2022- Q1/2023',
        template="simple_white",
        labels={ "Actual Contractual Commitment ($)": "Actual Commitment ($)"}
      )
```

FUND BY AREAS IN THE PERIOD Q1/2022- Q1/2023
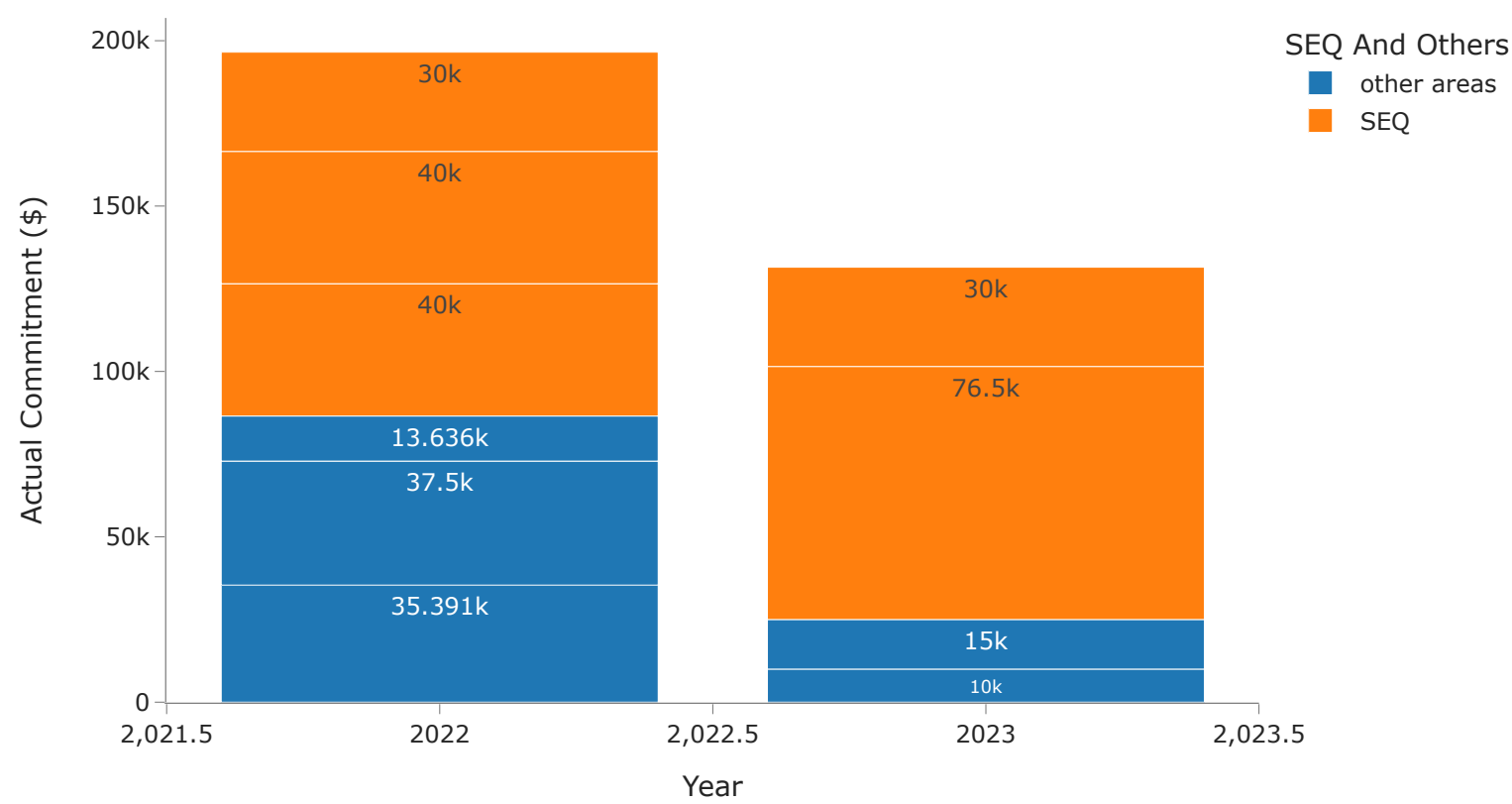
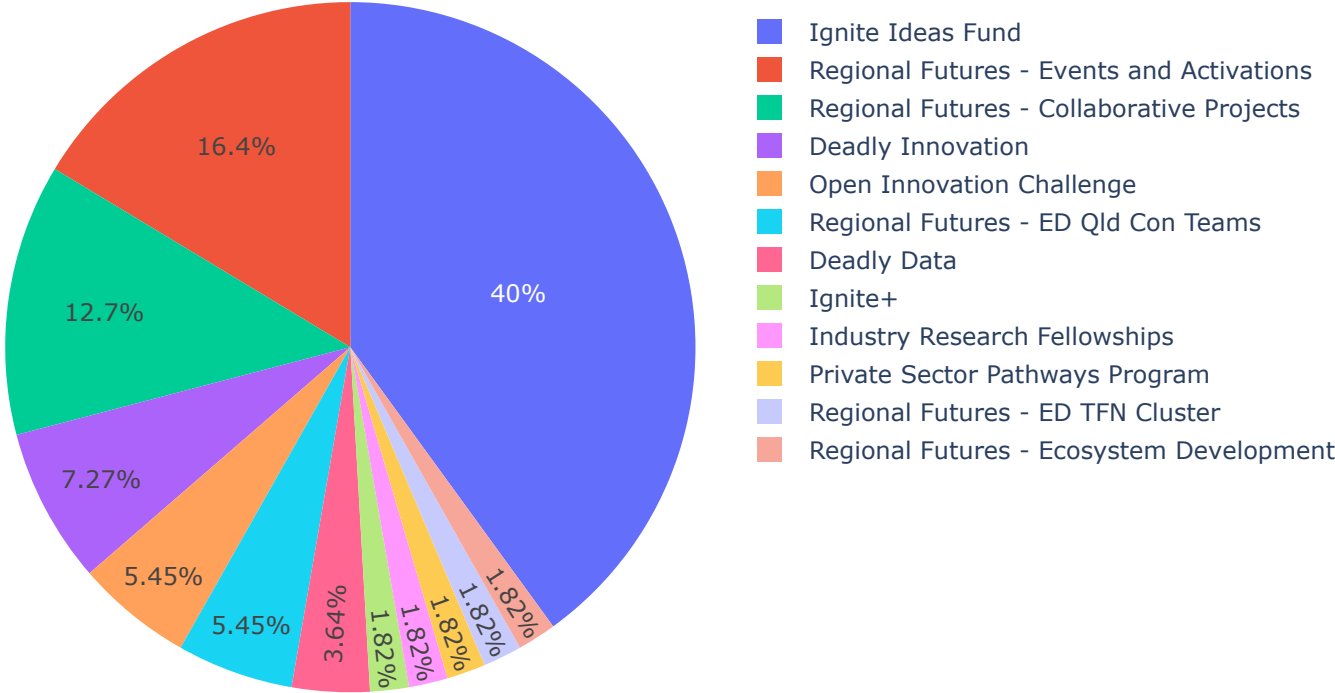### Insights:

- Because of limited of the limited dataset only has data in Jan, Feb, Mar of 2023, for better comparison, I chose data in the same period in 2022 to calculate
- Funding distribution was nearly the same in the observation period, and the gap between SEQ and other areas was also wider, while in 2023, the government decided to spend more for the SEQ area.

```
In [29]: #Data only 2022:
         df_2022 = df[(df['Approval date'].dt.year == 2022)]
         df_2022_program = df_2022.groupby('Program')['Actual Contractual Commitment ($)'].sum()
         #Program name
         #Create a chart by total:
         px.pie(df_2022, names = 'Program', title = 'DISTRIBUTION OF PROGRAM IN 2022')
         ### Write comment about the change from the name of Program between 2022 and 2023 strategies.
```

DISTRIBUTION OF PROGRAM IN 2022

Legend:
- Ignite Ideas Fund
- Regional Futures - Events and Activations
- Regional Futures - Collaborative Projects
- Deadly Innovation
- Open Innovation Challenge
- Regional Futures - ED Qld Con Teams
- Deadly Data
- Ignite+
- Industry Research Fellowships
- Private Sector Pathways Program
- Regional Futures - ED TFN Cluster
- Regional Futures - Ecosystem Development

```
In [30]: #Data only 2023:
         df_2023 = df[df['Approval date'].dt.year == 2023]
         #Create a chart by total:
         px.pie(df_2023, names = 'Program', title = 'DISTRIBUTION OF PROGRAM IN 2023', values = 'Actual Contractual Co
```

DISTRIBUTION OF PROGRAM IN 2023

Legend:
- Female Founders Program
- Regional Futures - Events and Activations
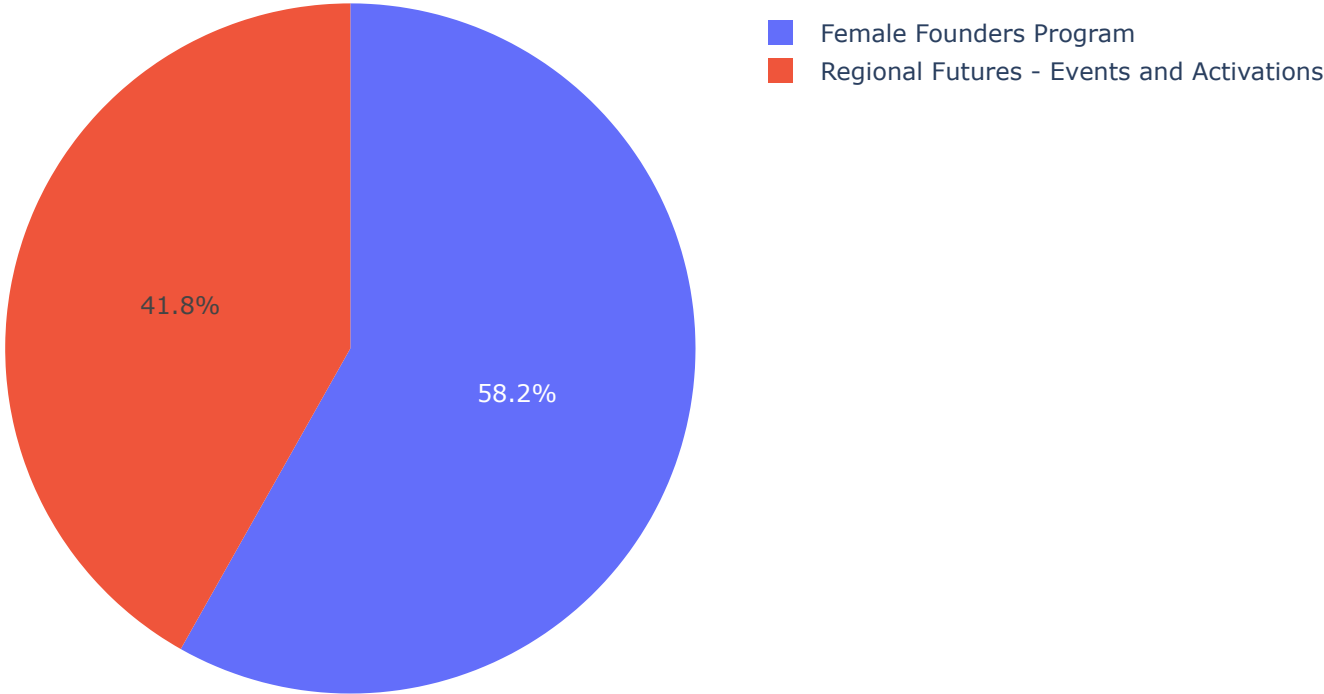
In [31]:
```python
# Distribution of program in 2023
regional = df_2023.groupby(['Program','Primary Location of Activity/Project - Suburb'])['Actual Contractual C
regional=regional.reset_index()
px.pie(regional,
       names = 'Primary Location of Activity/Project - Suburb',
       values = 'Actual Contractual Commitment ($)',
     title = 'DISTRIBUTION OF PROGRAM IN 2023 BY SUBURB')
```

DISTRIBUTION OF PROGRAM IN 2023 BY SUBURB

- Brisbane City
- South Brisbane
- Weipa
- Bundaberg

58.2%
22.8%
11.4%
7.6%
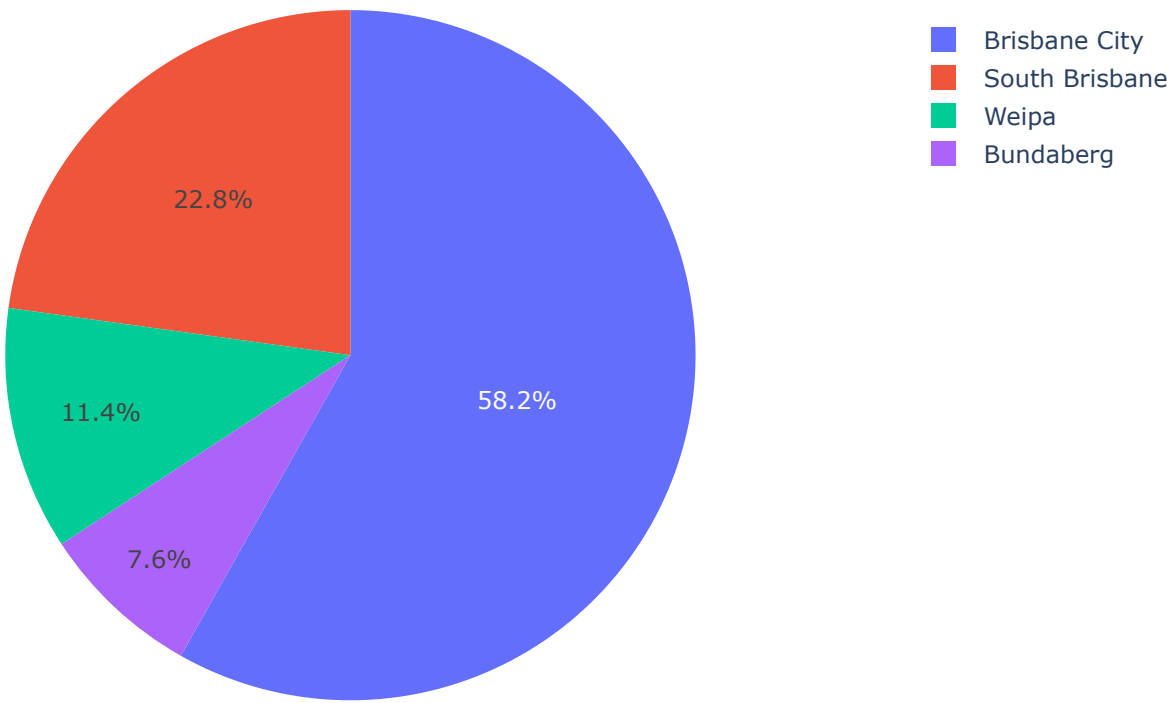
***Insights*:**

- But when we take a closer look, there is positive insight that the government has taken action on the region by spending funds to invest in Weipa and Bundaberg through the 'Regional Future' fund.
- The other thing is that they are promoting gender equality by using the Female Founders Program to invest in projects for women.
- In 2022, the Queensland government spent approximately 30% funds on regional (16,4% for 'Regional Future - Event and Activations' ; 12,7% on 'Regional Future - Collaborative Projects'). When in 2023, they spent nearly 42% on regional areas

CONCLUSION

Queensland Advance's funding situation has been decreasing since 2018, but the decline became more severe in 2021, 2022, and 2023 due to the impact of the COVID situation. Given the situation's lack of funds over the years and the worldwide downtrend of economics, the government must have a more efficient funding scheme than in the past. This report has been conducted to see if their funds are being used effectively. Also, the unequal fund spending between Sounthest Queensland and other areas and the support for specific community groups by funding need to be concerned. After analysis, some results have been provided:

- The amount of funds by Queensland Advance tends to decrease in the period 2017 - 2023
- Queensland government strategy in this period was focused on Southeast Queensland because it is an important area. This area has the highest population density and can be considered the head of the whole state in economic aspect.
- Queensland Advance's strategy will likely spend funds on research and education over the years. While three of the five most considerable funds are related to research ('Industry Research Fellowships', 'Research Fellowships, and 'Platform Technology Program')
- On the other hand, the 'Ignite Ideas Fund' for small and medium businesses has been the biggest fund over the observation period. This action shows that Queensland's government also considers economic growth in addition to research and education.
- But the limitation is that they do not focus on regional and mental health fields, while regionals have massive potential because of their extensive area, and mental health is essential for citizens' well-being.
- In 2022, the Queensland government spent approximately 30% on regional (16,4% for 'Regional Future - Event and Activations' ; 12,7% on 'Regional Future - Collaborative Projects')
- In 2023, they spent nearly 42% on regional areas, which is a positive trend. It's so that the state's strategy is slowly changing from SEQ to regional areas.

REFERENCES

For coding:

I using content of [Geekforgeeks (https://www.geeksforgeeks.org/python-maximum-minimum-elements-position-list/)](https://www.geeksforgeeks.org/python-maximum-minimum-elements-position-list/) to figure out how to find maximum number in a list by using id.max(). I need it to find which is the biggest fund every year.

For visualisation:

I using the document of [Plotly Pie Chart (https://plotly.com/python/pie-charts/)](https://plotly.com/python/pie-charts/), [Plotly Bar Chart (https://plotly.com/python/bar-charts/)](https://plotly.com/python/bar-charts/), and [Plotly Histogram (https://plotly.com/python/histograms/)](https://plotly.com/python/histograms/) to visualise my work.

For analysis:

I using the figure from [SEQ Councils (https://seqmayors.qld.gov.au/about-us/councils)](https://seqmayors.qld.gov.au/about-us/councils) to choose which suburbs belong to Southeast Queensland

In [ ]:

In [ ]: