# Welcome to the Little Red Dot:

## Analysis of International Arrivals to Singapore Data

## Acknowledgements & Thanks

This project would not have been possible without the help and support of my partner, Jocelyn. Thank you for taking the time to vet my codes, slides and understanding of regression analysis. Your comments and personal efforts made significant improvements to the final product.

Any mistakes in this project are purely my own.

## Project Objectives

“Welcome to the Little Red Dot” is a project examining data on international arrivals to Singapore and uncovering key insights.

I embarked on this project to apply my knowledge in conducting data analysis with Python 3 (and a bit of R). Practicing these skills helped build greater familiarity and hone new skills in cleaning, analysing, visualising and interpreting data.

I hope you enjoy reading the project deck. Thank you for your time! 😄

# Executive Summary

Tourism is an important pillar of Singapore's economy and currently contributes around 4% of its GDP. The industry is dependent on the volume of international arrivals and tourist expenditures to perform well. This project aims to determine salient trends and factors driving international arrivals to Singapore by analysing relevant data from 2000-2018. We will perform data visualisation of and investigate general trends, as well as conduct regression analysis to identify significant factors influencing total international arrival numbers. These findings are translated into actionable insights to inform the strategies of relevant stakeholders.

In recent years, China, Indonesia and India have emerged as the top contributors of international arrivals to Singapore, presenting opportunities for the tourism industry. The steady growth of visitors from these countries has contributed to a general increase in international arrivals between 2000-2018. However, certain years experienced falls due to adverse events, such as SARS (2003), the Global Financial Crisis (2009) and MH370 (2014). Finally, international arrivals seem to follow a pattern each year. International arrivals usually peak in July, plunge in September, before surging again in December.

Our regression analyses confirm at least three statistically significant factors influencing the total number of international arrivals to Singapore. As total population and GDP per capita in a country rises, we expect to see an increase in international arrivals from them. Conversely, as a country gets further away from Singapore, international arrivals fall. Interestingly, we do not find a significant relationship between sharing a common language and international arrivals.

## Contents

1

# Project introduction

Context, Problem Scoping and Datasets

## 1 Data Sets (1)

This project uses four main datasets to conduct an analysis of international arrivals trends to Singapore between 2000-2018

The first file, "sg_intl_arrivals.csv", is part of a larger data set from the Department of Statistics (Singapore). This data set contains international arrival numbers per month from each country, between 1978 to 2019 (Jun). It was last retrieved on 12 August 2019 from:

https://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=1991

## 1 Data Sets (2)

The second file, "gdp_per_capita.csv", is part of a larger data set from the World Bank Group. This dataset contains annual GDP per capita (in constant 2010 US$) of 264 countries/territories/groups, between 1960 to 2018. It was last retrieved on 12 August 2019 from: https://data.worldbank.org/indicator/NY.GDP.PCAP.KD

The third file, "country_population.csv" is part of a larger dataset from the World Bank Group. This dataset contains annual population totals of 264 countries/territories/groups, between 1960 to 2018. It was last retrieved on 12 August 2019 from: https://data.worldbank.org/indicator/SP.POP.TOTL?view=chart

## 1 Data Sets (3)

The fourth file, "sgp_dist_lang.csv", is part of a larger dataset from the CEPII. This data set provides bilateral data, such as distance between countries (in km), and dummy variables indicating if two countries share a common language or a colonial relationship.

We are interested in:

- Common language: whether another country shares a spoken language with Singapore (i.e. English, Chinese, Malay)
- Distance (km) between a particular country and Singapore

The data sets and accompanying notes were last retrieved on 12 August 2019 from: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6

# 1 Introduction

- Direct contribution of the travel & tourism sector to Singapore's GDP was $17.7 billion
  - Represents 4.1% of total GDP in 2017
- Tourism also serves to reinforce Singapore's status as a vibrant global city
- Crucial to understand trends and factors influencing international arrivals to the country
- Insights drawn can help stakeholders (e.g. govt., retailers, hotel management) capitalise on emerging trends

# 1 Problem Scoping

- Analysis focuses on the period between 2000-2018
- Part One: General Trends
  1. What is the general trend of international arrivals to Singapore? Are there reasons for a fall in certain years?
  2. Where are the bulk of international arrivals from?
  3. Are there 'cyclical effects' on international arrivals each year? Why?
- Part Two: Influential Factors
  1. What is the effect of GDP per capita on international arrivals?
  2. What is the effect of total population on international arrivals?
  3. Does sharing a common language increase international arrivals?
  4. How does geographic distance affect international arrivals?

# 1 Methodology

- Part One: General Trends
  a. Conduct general trend analysis and visualise data → further investigate any anomalous observations

- Part Two: Influential Factors
  a. Perform a cross-sectional regression analysis on 2018 international arrivals
  b. Perform a fixed effects analysis across 2000-2018

2

# Data import & exploration

Structure & Characteristics of dataset

## 2 Data Cleaning

- Prior data cleaning and merging was conducted (codes can be found in 'Data Cleaning and Merging.ipynb')

- Four datasets were processed and merged to form a single set:
  1. International arrivals to Singapore (DOS Singapore)
  2. Annual GDP per capita (2010 US$) (World Bank)
  3. Annual Country Population (World Bank)
  4. Weighted Distance and Official Language (CEPII)

# 2 Synopsis of merged dataset

| | country | year | visitor_count | country_code | total_pop | gdp_per_capita | comm_lang | distw_km |
|---|---|---|---|---|---|---|---|---|
| 0 | Australia | 2000 | 510,347.000 | AUS | 19,153,000.000 | 44,313.318 | 1 | 5,893.497 |
| 1 | Australia | 2001 | 550,681.000 | AUS | 19,413,000.000 | 44,564.977 | 1 | 5,893.497 |
| 2 | Australia | 2002 | 538,378.000 | AUS | 19,651,400.000 | 45,786.643 | 1 | 5,893.497 |
| 3 | Australia | 2003 | 392,891.000 | AUS | 19,895,400.000 | 46,575.415 | 1 | 5,893.497 |
| 4 | Australia | 2004 | 561,219.000 | AUS | 20,127,400.000 | 47,880.612 | 1 | 5,893.497 |
| 5 | Australia | 2005 | 620,196.000 | AUS | 20,394,800.000 | 48,760.355 | 1 | 5,893.497 |
| 6 | Australia | 2006 | 691,632.000 | AUS | 20,697,900.000 | 49,408.053 | 1 | 5,893.497 |
| 7 | Australia | 2007 | 768,490.000 | AUS | 20,827,600.000 | 50,955.056 | 1 | 5,893.497 |
| 8 | Australia | 2008 | 833,156.000 | AUS | 21,249,200.000 | 51,770.907 | 1 | 5,893.497 |
| 9 | Australia | 2009 | 830,299.000 | AUS | 21,691,700.000 | 51,689.914 | 1 | 5,893.497 |

## 2 Structure of data set

- Annual arrivals data (2000-2018) from various countries
  - E.g. Malaysia, Australia, China, India, UK, USA
  - 40 countries/markets represented in the dataset
- Information such as annual GDP per capita, total population, common official language and distance (from Singapore) are also included
- 760 rows, 8 columns

## 2 Summary Statistics

| | visitor_count | year | gdp_per_capita | total_pop | comm_lang | distw_km |
|---|---|---|---|---|---|---|
| **count** | 969.000 | 969.000 | 760.000 | 760.000 | 760.000 | 760.000 |
| **mean** | 223,888.077 | 2,009.000 | 28,191.156 | 118,922,447.446 | 0.500 | 6,845.159 |
| **std** | 444,763.643 | 5.480 | 23,128.496 | 273,207,324.289 | 0.500 | 3,813.456 |
| **min** | 0.000 | 2,000.000 | 346.775 | 333,241.000 | 0.000 | 505.538 |
| **25%** | 7,621.000 | 2,004.000 | 4,744.173 | 7,338,400.750 | 0.000 | 3,525.259 |
| **50%** | 42,415.000 | 2,009.000 | 31,438.436 | 46,624,019.500 | 0.500 | 6,825.147 |
| **75%** | 252,433.000 | 2,014.000 | 45,562.808 | 83,672,660.000 | 1.000 | 10,084.225 |
| **max** | 3,416,475.000 | 2,018.000 | 92,121.421 | 1,392,730,000.000 | 1.000 | 15,122.620 |

*The large ‘max’ values for visitor_count and total_pop can be attributed to China

3

# General trends

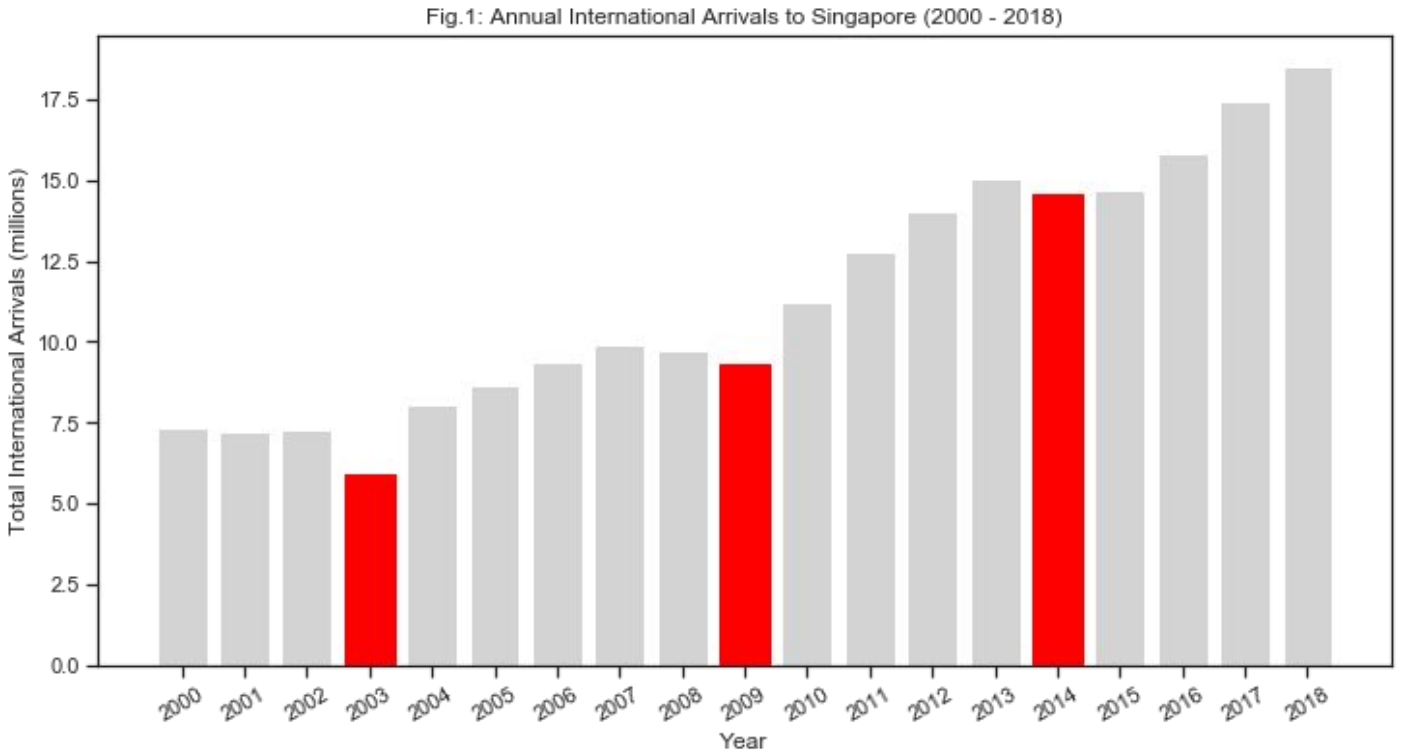
General insights and visualisations of the data

3.1

# Annual Arrivals

What is the general trend of international arrivals to Singapore? Are there reasons for a fall in certain years?

# 3.1 Annual arrival trends (2000-2018)



## 3.1 Annual arrival trends (2000-2018)

Preliminary Observations:

- International arrivals to Singapore generally increased between 2000-2018

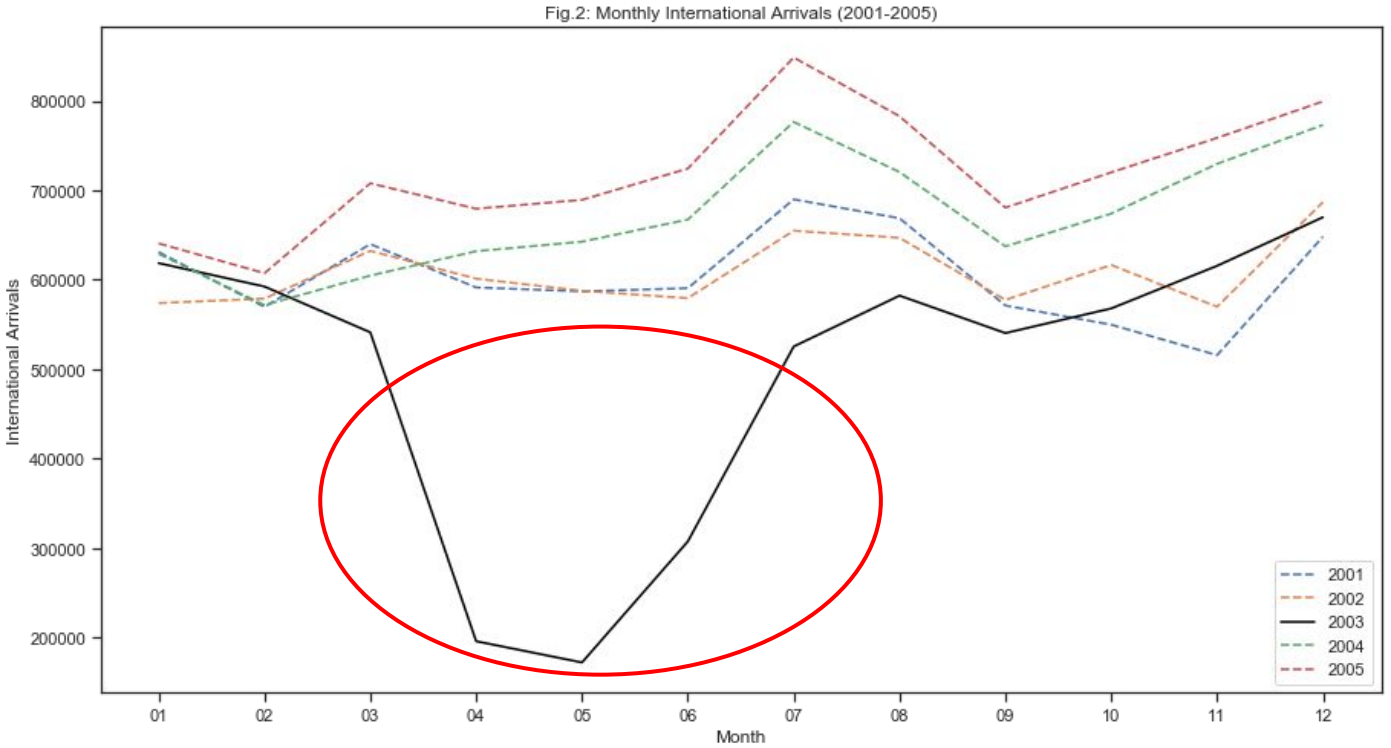
- However, there were certain years where total international arrivals fell - we turn towards investigating the reasons behind these dips (2003, 2009, 2014)

## 3.1.1 Why did arrivals fall in 2003?

- Hypothesis: The SARS epidemic in 2003 deterred foreigners from visiting Singapore to avoid the risk of being infected
- To affirm this hypothesis, we will visualise the monthly arrivals to Singapore between 2001 and 2005
- Patient No.1 was admitted to the hospital in March 2003 and the disease was eradicated in July 2003
- We would expect the period between March 2003 to July 2003 to see a sharp fall in international arrivals, and a recovery after

3.1.1

# Why did arrivals fall in 2003?



## 3.1.1 Why did arrivals fall in 2003?

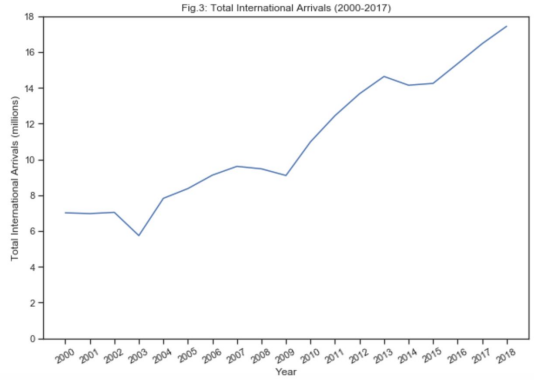
- The effect of SARS on international arrivals is clearly evidenced in the diagram – international arrivals plunged around March 2003 and only started recovering around June 2003
- Health scares or epidemics can adversely impact tourism in Singapore
- Singapore can mitigate such effects by ensuring its healthcare system is well-equipped to swiftly deal with health crises

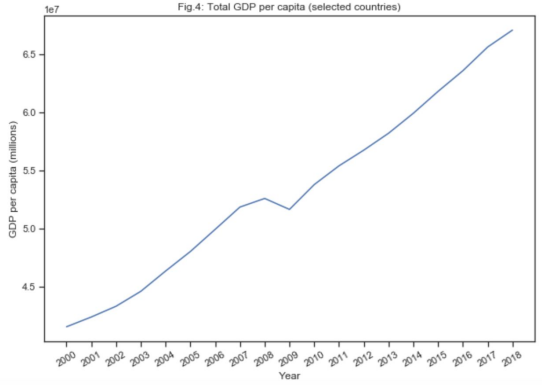
## 3.1.2 Why did arrivals fall in 2009 & 2014?

- Hypothesis: Dips in 2009 & 2014 were due to economic events - the Global Financial Crisis and Eurozone Crisis respectively
  - These economic crises likely reduced the disposable incomes of visitors and discouraged some from travelling to Singapore
- We will compare the overall trend of international arrivals with total GDP per capita of visiting countries over the same period
- The total GDP per capita is calculated by summing up the (non-null) total GDP the countries in the merged data set in each year, then divided by the total population of the countries in each year
  - This provides a gauge of the overall economic situation in the foreign countries present within the data set

3.1.2

# Why did arrivals fall in 2009 & 2014?




- When total GDP per capita of Singapore's guests fell in 2009, there was a corresponding (though less pronounced) dip in total arrivals

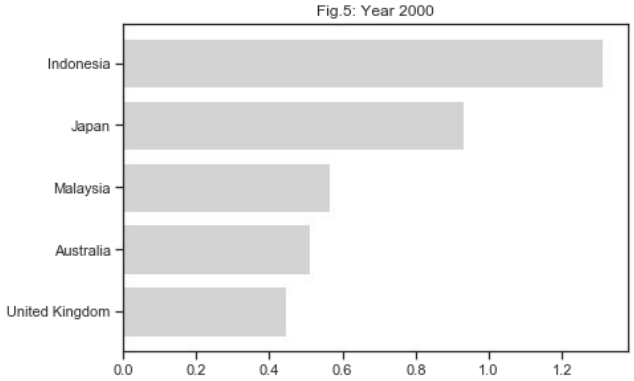
- However, the discernible fall in arrivals in 2014 seemed to occur when total GDP per capita of guest countries were rising - perhaps there were other factors which might better account for the decrease (see later section)

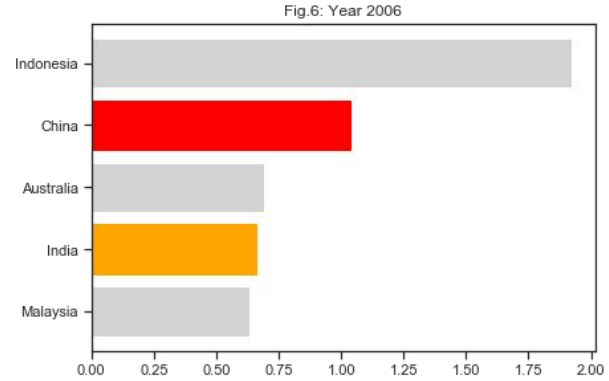
3.2

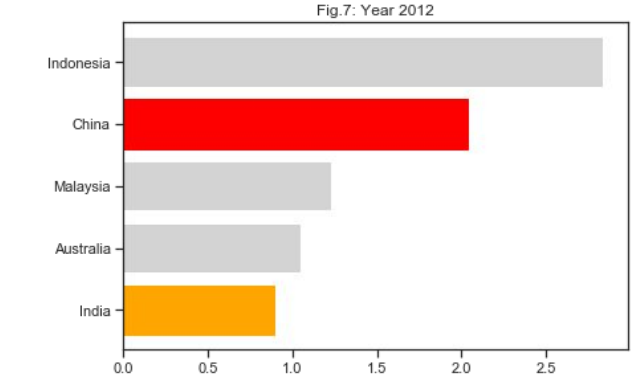
# Top countries by arrivals

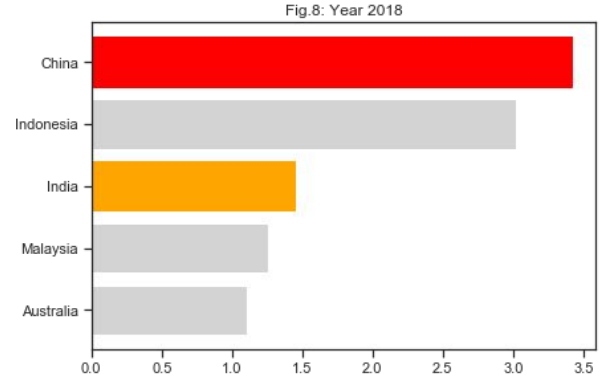
Where are the bulk of our visitors coming from?
How have these numbers evolved over time?

# 3.2 Top countries by arrivals (2000, 2006, 2012, 2018)



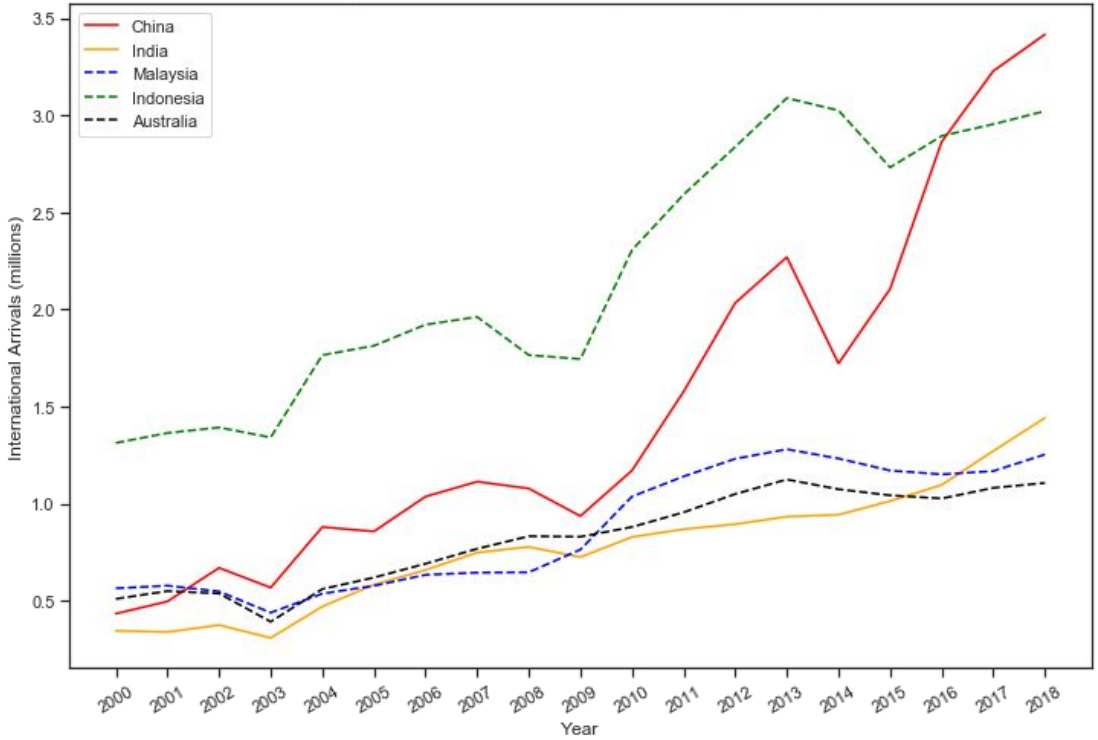



## 3.2.1 Arrival trends of top 5 countries

- From the snapshots, we observe that Malaysia, Indonesia and Australia have consistently been among the top 5 countries by international arrivals to Singapore
- More interestingly, China and India have broken into the top 5 – the number of international arrivals from these countries have grown significantly to even overtake other countries, with India in third place and China in first (in 2018)
- We next trace the developments in international arrival numbers from Australia, China, India, Indonesia and Malaysia between 2000–2018

# 3.2.1 Arrival trends of top 5 countries

Fig.9: International Arrivals from Top 5 Countries (2000-2018)



## 3.2.1 Arrival trends of top 5 countries

- Indonesia (dotted green) has been contributing a substantially greater amount of visitors relative to other countries between 2000-2018
- We see a steep rise in arrivals from China (e.g. after 2009 & 2014), even replacing Indonesia in 2016 as the largest source of visitors to Singapore
- Various stakeholders can capitalise on these emerging trends
    - Hotels can train their staff in Chinese etiquette and converse in Mandarin/Bahasa Indonesia to attend to their guests
    - STB is focusing marketing efforts in China, India and Indonesia to attract more visitors - this also serves to diversify the risk of over-relying on one country
    - Retailers are catching on with the surge in Chinese tourists - Suntec City and Alipay cemented a partnership to launch Alipay touchpoints in the mall, enabling a seamless shopping experience for Chinese visitors.

## 3.2.1 Arrival trends of top 5 countries (2)

- There are several plausible reasons accounting for the majority of visitors originating from these countries
  (1) Growth in population size
  (2) Rising affluence of the middle class in emerging economies
  (3) Geographical proximity
  (4) Cultural/linguistic similarities

- We will explore the relationship between these factors and international arrivals in greater detail later

## 3.2.2 ‘Mystery’ of 2014 fall (partly) solved?

- Referring to Fig.9, we note a precipitous fall in arrivals from China in 2014 – this sharp fall was not observed in the other four countries

- This plunge likely contributed to the overall decrease in international arrivals to Singapore in 2014

- This fall could be linked to the MH370 aviation disaster, where a plane from Kuala Lumpur to Beijing disappeared mid-flight
  - This may have scared Chinese tourists from visiting the region temporarily

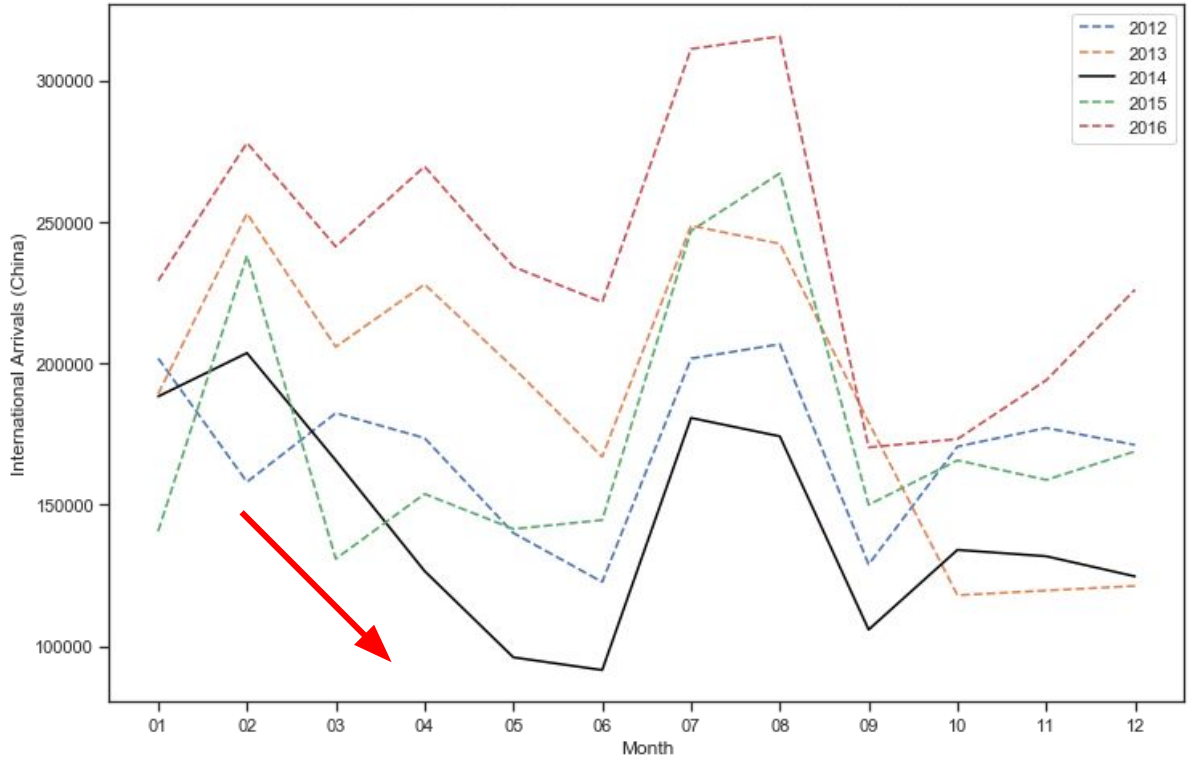
## 3.2.2 ‘Mystery’ of 2014 fall (partly) solved?

- To determine the effects of MH370 on Chinese arrivals in 2014, we will conduct an analysis similar to Section 3.1.1 (SARS)

- We will plot the trends of international arrivals from China between 2012 and 2016 and observe if there was an anomalous drop around the time MH370 occurred (March 2014)

# 3.2.2 ‘Mystery’ of 2014 fall (partly) solved?

Fig.10: Monthly International Arrivals from China (2012-2016)



## 3.2.2 ‘Mystery’ of 2014 fall (partly) solved?

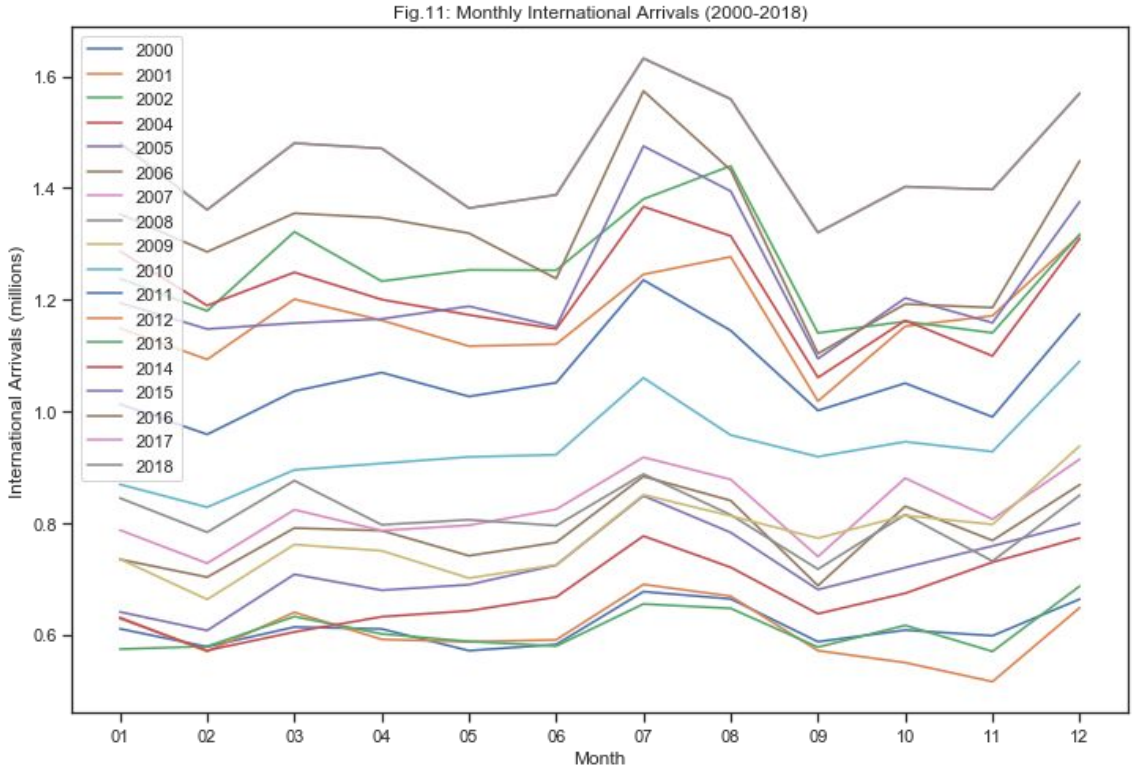
- The effect of MH370 is rather conclusive in Fig.10 – we see a sharp drop from Feb 2014 which lasted till June 2014
- This coincided with the period when the disaster occurred and was still fresh in people’s minds
- Even though arrival numbers rebounded in July 2014, it was still below those of 2012
    - This could indicate that MH370 probably still weighed on the Chinese's decision to travel to Singapore months after

3.3

# Annual seasonal effects

Are there ‘cyclical effects’ on international arrivals to Singapore each year? Why?

# 3.3 Annual seasonal effects



## 3.3 Annual seasonal effects

- Fig.11 reveals a few cyclical patterns in arrivals to Singapore between 2000-2018
  1. International arrival numbers usually peak in July each year
  2. There is a sharp fall in international arrivals in September every year
  3. International arrival numbers pick up and surge again in December (comparable to July's level) each year
- Preliminary desk research was conducted to unearth some reasons driving this consistent cyclical pattern

## 3.3 Annual seasonal effects

- Peak in July
  - Great SG Sale and SG Food Festival are being held during July
  - Coincides with the start of summer vacations in many countries
- Slump in September
  - Coincides with the end of summer vacation (early September)
  - Strange given that the F1 night race occurs during this month - would not expect such a sharp fall (after 2008)
- Surge in December
  - May be due to the influx of winter travellers
  - Zoukout, a weekend-long party, may also be a huge draw

## 3.3 Annual seasonal effects: Extension

- Desk research should be seen as a starting point and not conclusive towards explaining the cyclical trends in international arrivals

- To supplement our analysis, we could obtain demographic data of international visitors, such as their age & employment status (e.g. students would have seasonal holidays rather than working adults)

- Qualitative methods (e.g. interviews) can help uncover the reasons motivating foreigners to visit Singapore (or not) during specific months

## 3 General trends - Summary

- Annual international arrivals have generally increased between 2000-2018, though certain years experienced falls
  - Fall in 2003 was due to the SARS epidemic
  - Fall in 2009 was due to the Global Financial Crisis
  - Fall in 2014 could be partially due to the MH370 aviation disaster
- China, Indonesia and India are the top countries contributing to international arrivals in Singapore - Visitor numbers from China and India in particular are growing steadily and consistently
- There is a cyclical pattern in arrivals every year: arrival numbers peak in July, slump in September, and surge again in December - this could be due to seasonal effects such as summer holidays and winter travellers

4

# Influential factors

Regression analysis of potential factors accounting for international arrivals

# 4 Influential factors: Synopsis

- We investigate the following possible factors influencing international arrivals to Singapore through regression analysis:
  1. What is the effect of GDP per capita on international arrivals?
  2. What is the effect of total population on international arrivals?
  3. Does having a common language increase international arrivals?
  4. How does geographic distance affect international arrivals?
- We begin with exploratory data analysis after cleaning and transformation
- As the final dataset is a panel dataset that is changing across countries and time, we will have to conduct both a cross-sectional analysis for factors that are changing across countries, and a fixed-effects analysis for factors that are changing across time.

# 4.1 Data cleaning & EDA

- Clean the merged dataset by removing rows where ‘visitor_count’ value is 0 and ‘country_code’ is ‘None’
- Removed all entries in 2003 as it was an anomalous year
- 624 entries in total, first 10 rows below:

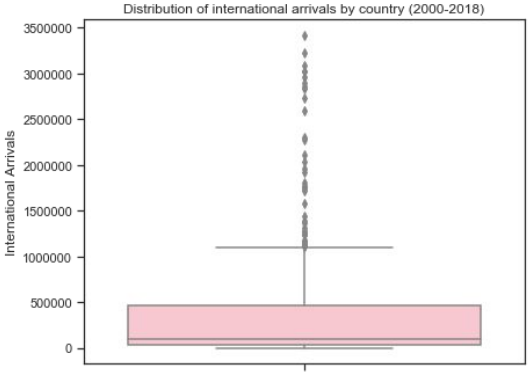
| country_code | country | year | visitor_count | total_pop | gdp_per_capita | comm_lang | distw_km |
|---|---|---|---|---|---|---|---|
| AUS | Australia | 2000 | 510,347.000 | 19,153,000.000 | 44,313.318 | 1 | 5,893.497 |
| AUS | Australia | 2001 | 550,681.000 | 19,413,000.000 | 44,564.977 | 1 | 5,893.497 |
| AUS | Australia | 2002 | 538,378.000 | 19,651,400.000 | 45,786.643 | 1 | 5,893.497 |
| AUS | Australia | 2004 | 561,219.000 | 20,127,400.000 | 47,880.612 | 1 | 5,893.497 |
| AUS | Australia | 2005 | 620,196.000 | 20,394,800.000 | 48,760.355 | 1 | 5,893.497 |
| AUS | Australia | 2006 | 691,632.000 | 20,697,900.000 | 49,408.053 | 1 | 5,893.497 |
| AUS | Australia | 2007 | 768,490.000 | 20,827,600.000 | 50,955.056 | 1 | 5,893.497 |
| AUS | Australia | 2008 | 833,156.000 | 21,249,200.000 | 51,770.907 | 1 | 5,893.497 |
| AUS | Australia | 2009 | 830,299.000 | 21,691,700.000 | 51,689.914 | 1 | 5,893.497 |
| AUS | Australia | 2010 | 880,558.000 | 22,031,750.000 | 51,936.889 | 1 | 5,893.497 |

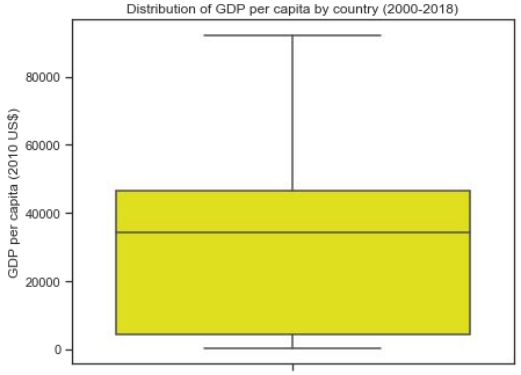
## 4.1 Data cleaning & EDA

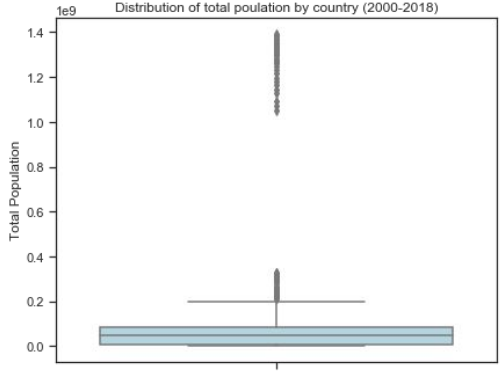
- Visualise the distribution of international arrivals (red), GDP per capita (yellow), total population (blue) and weighted distance from Singapore (green) using boxplots

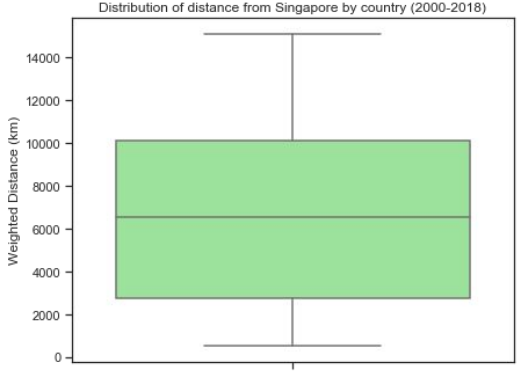
- Obtain summary statistics (e.g. percentiles, mean) and outliers

# 4.1 Data cleaning & EDA



## 4.1 Data cleaning & EDA

- The distribution of international arrivals and total population contain many outliers that can severely affect our analysis
- Outliers in international arrivals are caused by China, Indonesia and India, each with over a million arrivals to Singapore in recent years
- Outliers in total population are due to China and India, with population sizes of over 1 billion people
- The presence of outliers highlights the need for additional steps to be taken before we begin our analysis

4.2

# Cross-sectional analysis

International arrivals data in 2018:
Which factors are statistically significant and what is their relationship with international arrivals?

## 4.2 Cross-sectional analysis

- Slice out a dataframe from the modelling dataset introduced in Section 4.1
- 39 rows/data points in total – first 5 rows displayed
- Year of analysis is restricted to 2018 to exploit the variation in the 4 factors across countries in 2018.

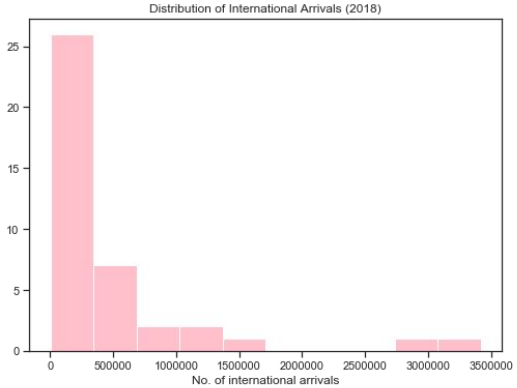
| country_code | country | year | visitor_count | total_pop | gdp_per_capita | comm_lang | distw_km | log_visitor_count | log_total_pop | log_gdp_per_capita |
|---|---|---|---|---|---|---|---|---|---|---|
| AUS | Australia | 2018 | 1107215 | 24992369 | 56919.374 | 1 | 5893.497 | 13.91736 | 17.03408 | 10.949391 |
| BGD | Bangladesh | 2018 | 126301 | 161356039 | 1203.216 | 0 | 2874.279 | 11.74642 | 18.89912 | 7.092753 |
| BRN | Brunei Darussalam | 2018 | 74960 | 428962 | 31436.949 | 1 | 1264.269 | 11.22471 | 12.96912 | 10.355739 |
| CAN | Canada | 2018 | 129512 | 37058856 | 51357.754 | 1 | 14394.100 | 11.77153 | 17.42802 | 10.846571 |
| CHN | China | 2018 | 3416475 | 1392730000 | 7754.962 | 1 | 4097.429 | 15.04412 | 21.05453 | 8.956088 |
| DNK | Denmark | 2018 | 37417 | 5797446 | 62888.727 | 0 | 10051.390 | 10.52988 | 15.57293 | 11.049122 |

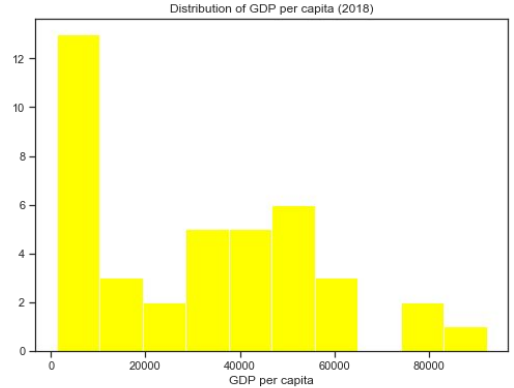
## 4.2 Cross-sectional analysis - Data transformation

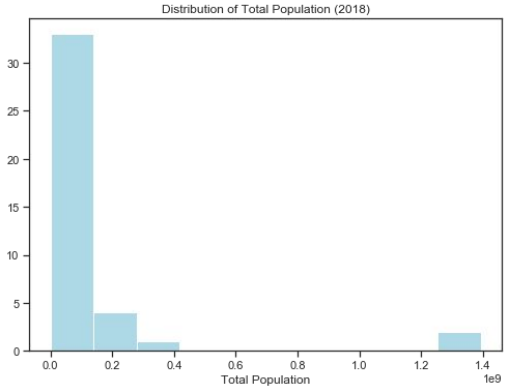
- Plot histograms of 2018 international arrivals (red), GDP per capita (yellow), total population (blue) and weighted distance from Singapore (green)

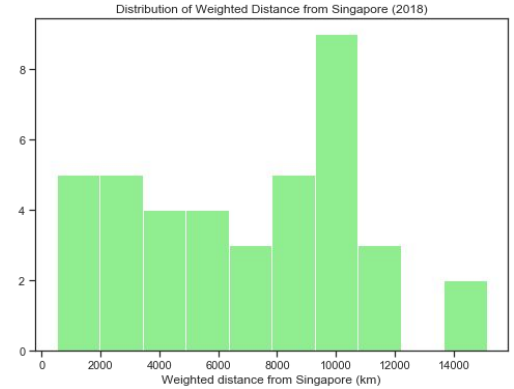
- Visualise the distribution of each variable – they should ideally be symmetric and normally distributed for the regression model to work better

# 4.2 Cross-sectional analysis - Data transformation



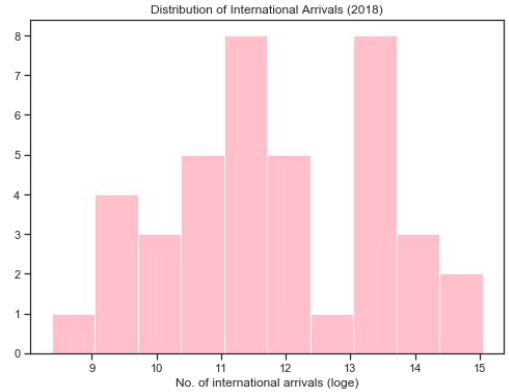
## 4.2 Cross-sectional analysis - Data transformation

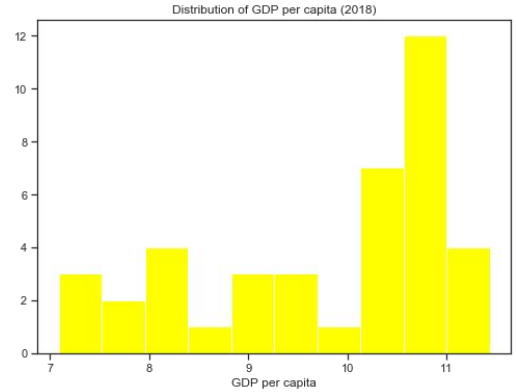
- The distributions of international arrivals and total population are asymmetrical and have huge outliers

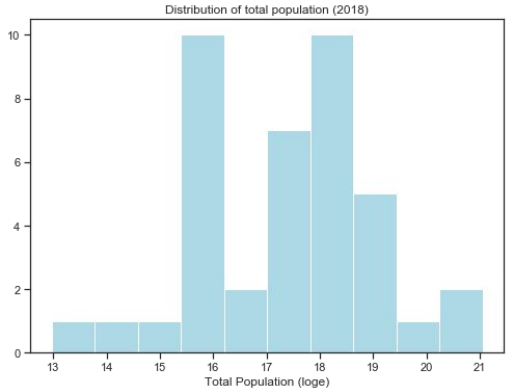
- We resolve this by performing a “natural log” transformation on these two variables to alter the shape of their distribution
  - This should improve the shape of their distributions and reduce the influence of outliers
  - We will also do a “natural log” transformation on GDP per capita to make the regression results more interpretable

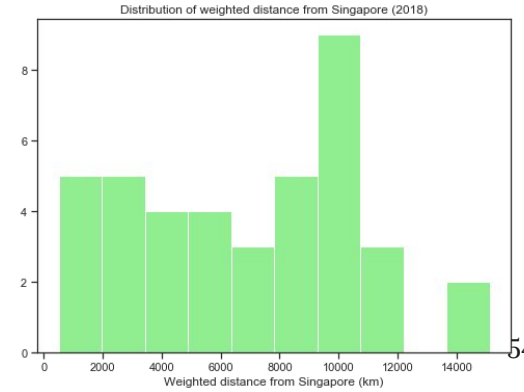
# 4.2 Cross-sectional analysis - Data transformation

"Natural log" transformation:
More symmetric distribution,
outliers are less pronounced



# 4.2 Cross-sectional analysis - Data transformation

2018 Cross-sectional analysis dataset after “natural log” transformation

| country_code | country | visitor_count | year | gdp_per_capita | total_pop | comm_lang | distw_km | log_visitor_count | log_total_pop | log_gdp_per_capita |
|---|---|---|---|---|---|---|---|---|---|---|
| AUS | Australia | 1107215 | 2018 | 56919.374 | 24992369 | 1 | 5893.497 | 13.91736 | 17.03408 | 10.949391 |
| BGD | Bangladesh | 126301 | 2018 | 1203.216 | 161356039 | 0 | 2874.279 | 11.74642 | 18.89912 | 7.092753 |
| BRN | Brunei Darussalam | 74960 | 2018 | 31436.949 | 428962 | 1 | 1264.269 | 11.22471 | 12.96912 | 10.355739 |
| CAN | Canada | 129512 | 2018 | 51357.754 | 37058856 | 1 | 14394.100 | 11.77153 | 17.42802 | 10.846571 |
| CHN | China | 3416475 | 2018 | 7754.962 | 1392730000 | 1 | 4097.429 | 15.04412 | 21.05453 | 8.956088 |
| DNK | Denmark | 37417 | 2018 | 62888.727 | 5797446 | 0 | 10051.390 | 10.52988 | 15.57293 | 11.049122 |

# 4.2 Cross-sectional analysis - Regression model

Stepwise regression was performed on R to produce the model on the right

**Estimated Model:**

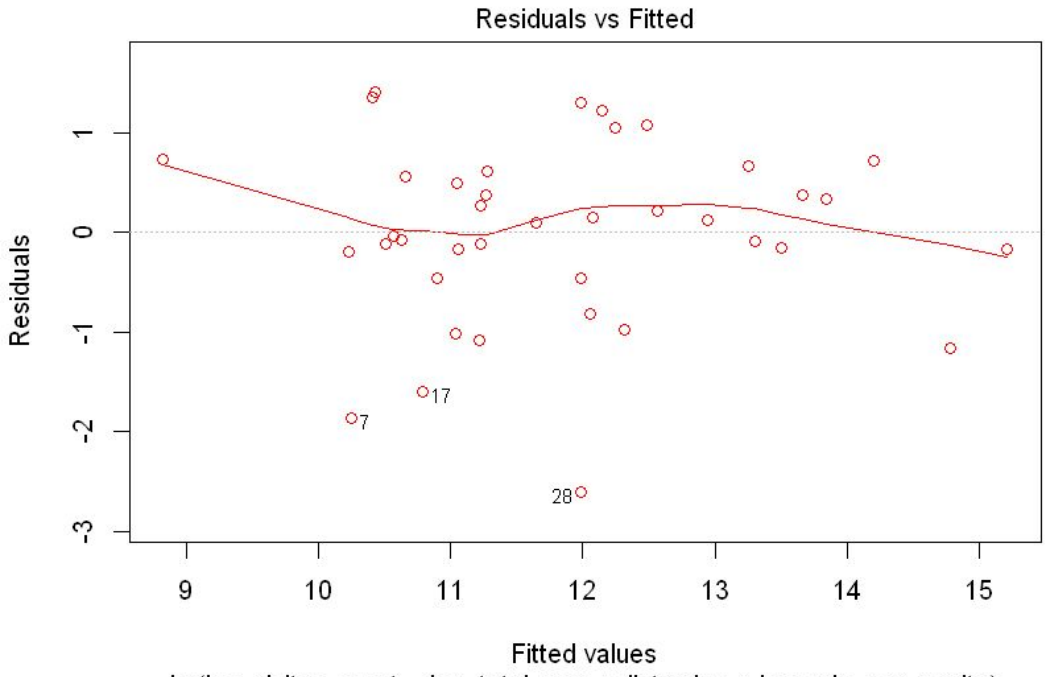
*log(Total Arrivals)* =
-12.05 + 0.886 log(Total Population) -0.362 Distance + 1.13 log(GDP per capita)

| Dependent variable: log(Total Arrivals) | |
|---|---|
| **log(Total Population)** | 0.8856*** (0.1099) |
| **Distance from Singapore ('000 km)** | -0.3621*** (0.0552) |
| **log(GDP per capita (2010 US$))** | 1.129*** (0.1887) |
| **Intercept** | -12.05*** |
| **Adjusted $R^2$** | 0.6778 |
| **No. of Obs.** | 39 |

## 4.2 Cross-sectional analysis - Interpreting results

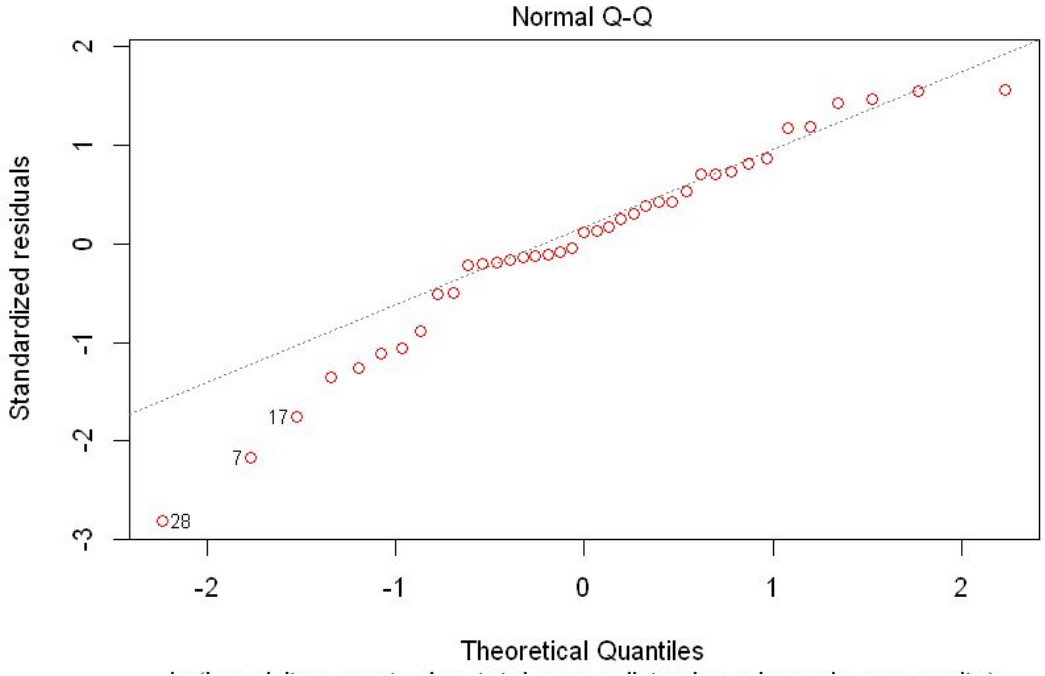
- Holding all else constant, if Total Population increases by 1%, we expect Total Arrivals to increase by 0.89%
- Holding all else constant, if Distance from Singapore increases by 1000 km, we expect Total Arrivals to decrease by 36%
- Holding all else constant, if GDP per capita increases by 1%, we expect Total Arrivals to increase by 1.13%
- There is evidence to infer that log(population), log(GDP per capita) and distance from Singapore are linearly-related to total arrivals - these coefficients are significant at the 1% level
- Conversely, there is insufficient evidence to conclude that sharing a common language is associated with arrival numbers from a foreign country, so it was dropped from the analysis
- The model can explain around 67.78% of the variation in international arrivals data

# 4.2 Cross-sectional analysis - Fitted vs. Residual plots



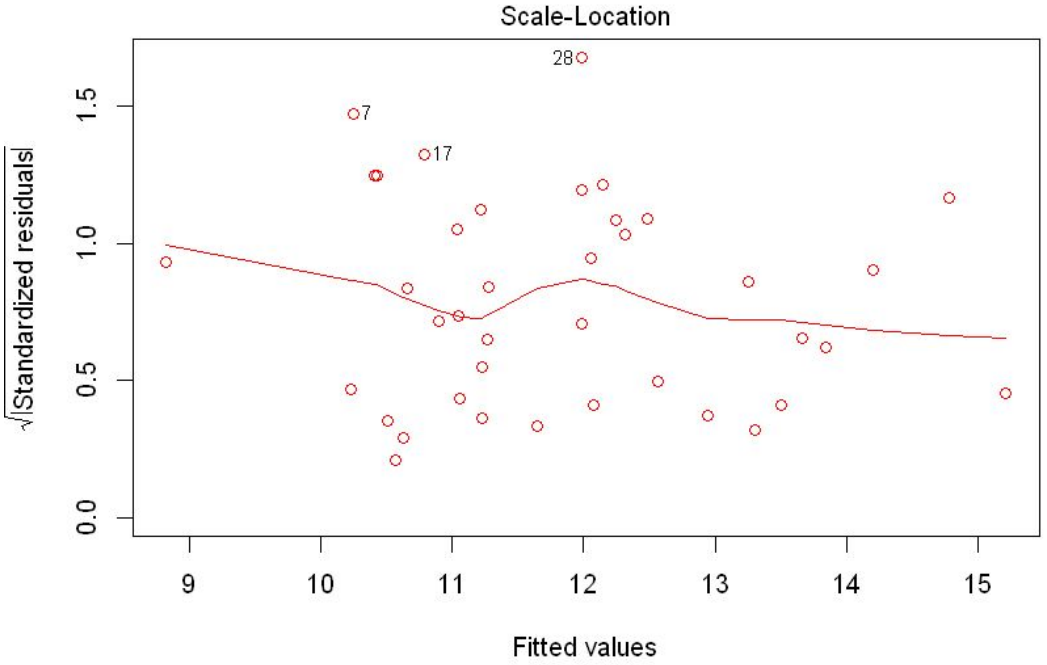

- The residuals seem to be randomly scattered around zero (no distinctive pattern), though there are a few residuals around -2
- The residuals do not seem to have non-linear patterns - a good indication that we do not have non-linear relationships among the variables
- The deterministic component of the model explains the dependent variable well, leaving only inexplicable random errors

# 4.2 Cross-sectional analysis - Q-Q plots



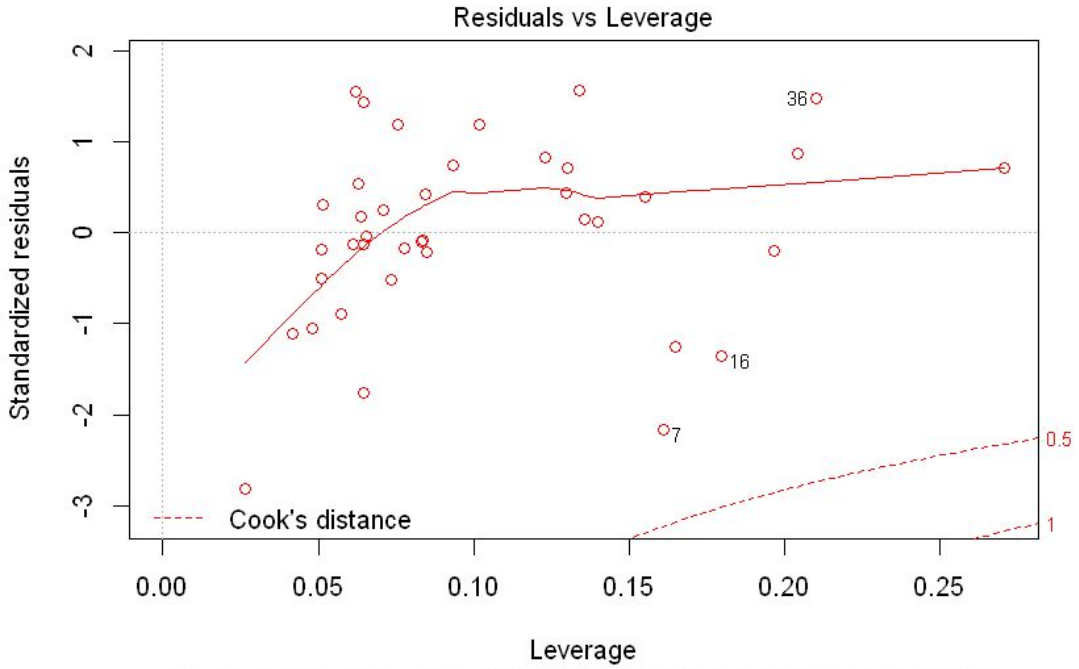

- Majority of the residuals follow rather closely to the straight line on the plot, which is an indicator that they are normally distributed
- The Q-Q plot indicates a slightly left-skewed distribution - however with $n$ = 39, the sample size may not be large enough to be conclusive
- Overall, for our model, the Q-Q plot shows reasonable alignment to the line, though we note that the numbered points are a bit off

# 4.2 Cross-sectional analysis - Spread-location plot




- The residuals are reasonably spread above and below a rather horizontal line
- We do note the bump in the middle of the line, perhaps due to the point no. 29 being far above it
- As we only have 39 data points, each residual has a more pronounced effect on the shape of the line as well

# 4.2 Cross-sectional analysis - Residuals vs. Leverage




- The plot does not reveal any influential cases that can significantly affect the regression model - none of the case are beyond the Cook's distance

## 4.2 Cross-sectional analysis - Summary

- The stepwise regression model of international arrivals can be considered valid based on the diagnostic plots we have conducted
- The model suggests that population, GDP per capita and distance from Singapore are statistically significant in influencing total international arrivals – their relationships also make intuitive sense
    - If a country's total population increases, we can expect more arrivals from them
    - If a country's GDP per capita increases, we can expect more arrivals from them
    - The further a country is from Singapore, the fewer international arrivals we can expect
- Interestingly, sharing a common language did not show up as a statistically significant factor contributing towards international arrivals

4.3

# Fixed effects analysis

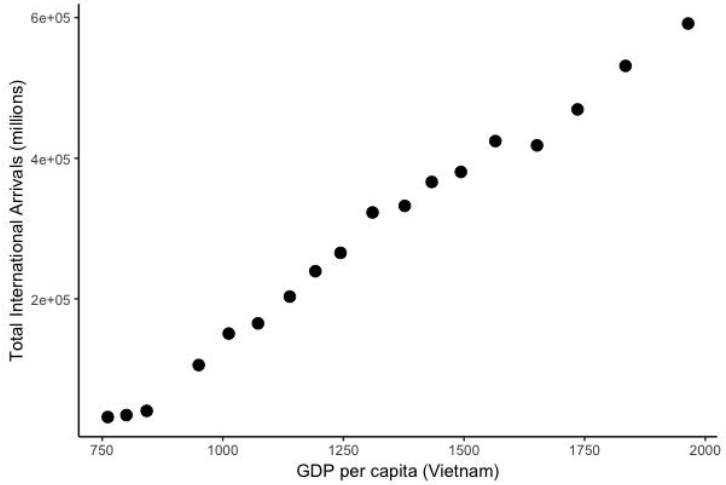
International arrivals data across time:
By controlling for each country, which factors are statistically significant and what is their relationship with international arrivals?

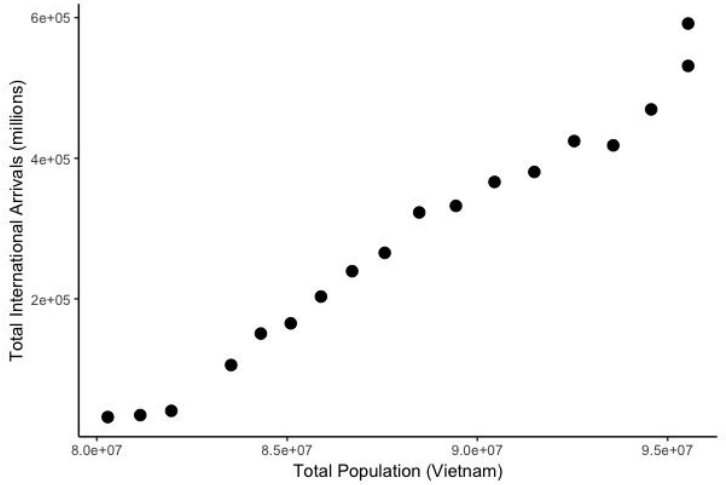
## 4.3 Fixed effects analysis

- Section 4.2 studied the variation across countries in 2018 only
- In this section, we examine the variation across time (2000–2018) within each country
- We address the following question: "For country X, as its GDP and population change across time from 2000–2018, how does its total visitors to Singapore change?"
- The fixed effects model asks this question about every country in the dataset and arrives at an overall conclusion about the relationship between GDP and population with visitor count
- The limitation of a fixed effects model is that we cannot look at factors that are not changing over time for every country
    - We hence have to rely on the previous cross-sectional analysis to explore the effects of distance from Singapore and common language since these are factors which do not change across time, but across country.

# 4.3 Justification for Fixed effects analysis

- If we zoom in to particular countries, the relationship between GDP per capita and Total Visitor Arrivals and between Total Population and Total Visitor Arrivals is very clear
- As such, a fixed effects model will help us sieve out these relationships.
- For instance, for Vietnam:



# 4.2 Fixed effects analysis - Regression Model

Fixed effects model run using R

**Estimated Model:**

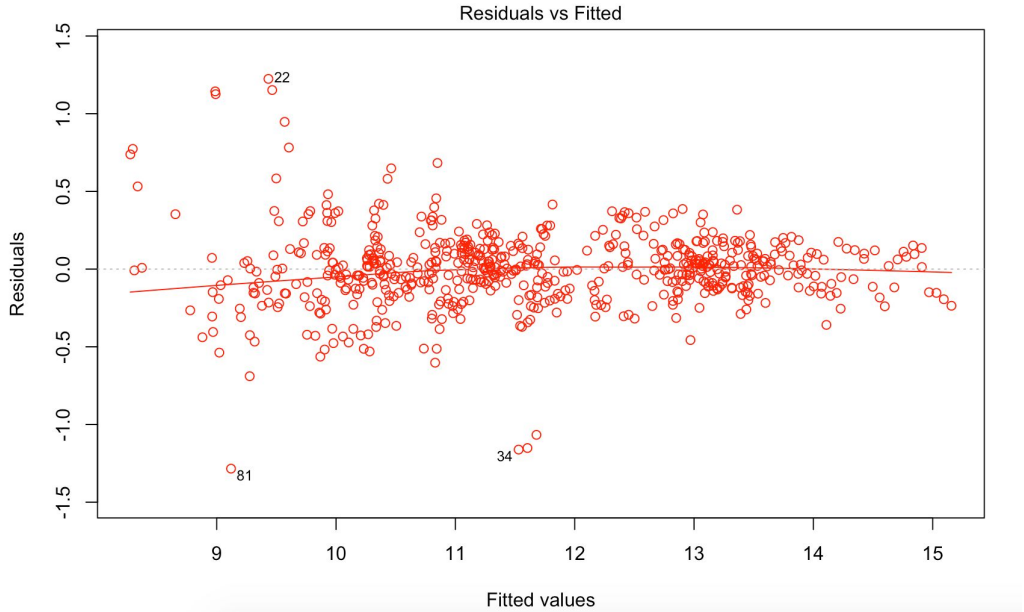
*log(Total Arrivals)* =
$\alpha$i + 1.58 log(Total Population) + 1.16 log(GDP per capita),
where $\alpha$i represents the country-fixed effect for country i

| Dependent variable: log(Total Arrivals) | |
|---|---|
| **log(Total Population)** | 1.58*** (0.105) |
| **log(GDP per capita (2010 US$))** | 1.16*** (0.06) |
| **Adjusted $R^2$** | 0.9722 |
| **No. of Obs.** | 623 |

## 4.2 Fixed effects analysis - Interpreting results

- Holding all else constant, if Total Population increases by 1%, we expect Total Arrivals to increase by 1.58% (Higher than in cross-section)
- Holding all else constant, if GDP per capita increases by 1%, we expect Total Arrivals to increase by 1.16% (similar to cross-section analysis)
- There is evidence to infer that log(population), log(GDP per capita) are linearly-related to total arrivals - these coefficients are significant at the 1% level
- The model can explain around 97.22% of the variation in international arrivals data
- We find that the effects of population is actually more pronounced when we use the fixed effects model compared to the cross-sectional analysis.
- This means that at the country level, we would expect growth in total population in a country to have a larger effect on total visitors from that country
- It is a positive sign that both cross-section and fixed effects corroborate each other in terms of the direction of the relationships

# 4.2 Fixed effects analysis - Fitted vs. Residual plots




- The residuals seem to be randomly scattered around zero
- The residuals do not seem to have non-linear patterns - a good indication that we do not have non-linear relationships among the variables
- The deterministic component of the model explains the dependent variable well, leaving only inexplicable random errors

5

# Conclusion & Reflections

Key takeaways, summary of insights and next steps

# 5.1 Conclusions - General trends

1. Singapore's tourism industry is vulnerable to external shocks, although measures can be taken to mitigate the detrimental effects
    - Diversify sources of international arrivals and not over-rely on one
    - Build resilience in other areas, such as healthcare infrastructure, domestic security and political stability to ensure Singapore remains a safe destination to visit
2. China, Indonesia and India have emerged as the top countries contributing to international arrivals to Singapore
    - Various stakeholders should adopt campaigns, technologies and training specifically targeting visitors from these regions to boost their performance metrics
3. International arrivals to Singapore generally follow a cyclical pattern each year - peak in July, slump in September and surge again in December
    - Insights on this cyclical pattern could help inform stakeholders' decisions in optimising resources (e.g. better plan staff allocation, room/stock availability and prices)

# 5.1 Conclusion - Influential factors

1. As total population and GDP per capita of a country grows, we can expect to see an increase in international arrivals from that country
    - Players in the tourism industry can consider designing and focusing their offerings to attract visitors from emerging economies in our region
2. Larger distances may be a prohibitive factor for international visitors (e.g. longer flight duration, costlier tickets or fewer available flights)
    - Improvements in air travel technologies in the future could potentially alter findings
    - This could also signal stakeholders to focus less on attracting visitors from countries further away, as the returns may not be commensurate
3. Sharing a common language does not seem to have any effect on international arrivals from the foreign country
    - It seems that linguistic similarities are not a major consideration when foreigners decide to visit Singapore

## 5.1 Conclusion - Caveats

- General trends analysis should be understood as supporting certain explanations and is by no means conclusive - more information can be collected to assess the proposed explanations
- Regression models are unable to account for all factors influencing international arrivals to Singapore
    - E.g. Availability of flights within each country
    - In the fixed effects model, the exact country-specific causes cannot be identified, but only controlled for
- Models may only be limited to Singapore's context and are not generalisable to other jurisdictions

## 5.1 Conclusion - Further analysis

- Qualitative studies and data (e.g. interviewing visitors) could supplement the existing analysis of the project
- Additional data of international arrivals, such as demographic information (age, gender) can extend our current trend analysis
- Explore other factors that may influence international arrivals to Singapore and incorporate into the regression analysis
- Quantify and assess the effects of STB marketing campaigns or high profile events on arrivals (e.g. Trump-Kim summit in Feb 2019) beyond anecdotal statements

## 5.2 Reflections - Personal learnings

- Experienced working on a full data analysis project: problem scoping → data collection & cleaning → data analysis & visualisation → interpreting & communicating results
- Practical application of Python and R coding knowledge to perform data analysis, visualisation and regression modelling
- Developing patience and troubleshooting skills (understanding the issue, then sourcing the internet and adapting code solutions)
- Opportunity to derive insights from data, structure findings and communicate actionables effectively
- Revised and better grasped the theories/applications of regression analysis
- Gained a deeper appreciation and understanding of the tourism industry in Singapore

6

# References

Resources & Credits

# 6 References


1. Changi Jewel photo by Joe Green on Unsplash
2. Brady, I. (2018, Mar 28). Residual Analysis in Linear Regression. Retrieved from https://rpubs.com/iabrady/residual-analysis
3. CEPII. (n.d.). *GEODIST* [Data file]. Retrieved from http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6
4. Dave, S. (2019, Feb 15). The best time to visit Singapore. [Blog Post]. Retrieved from https://santorinidave.com/best-time-to-visit-singapore
5. Department of Statistics Singapore. (2019). *M550001 - International Visitor Arrivals By Inbound Tourism Markets, Monthly* [Data file]. Retrieved from https://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=1991
6. Dragonfly Statistics. (2017, Oct 18). *Stepwise Regression in R - Combining forward and backward selection* [Video file]. Retrieved from https://www.youtube.com/watch?v=ejR8LnQziPY
7. Hong, J. (2018, Feb 12). China beats Indonesia as Singapore's top tourist source. *The Straits Times*. Retrieved from https://www.straitstimes.com/singapore/china-beats-indonesia-as-singapores-top-tourist-source
8. Kephart, C. (2013, Jun 13). Interpret Regression Coefficient Estimates. Retrieved from http://www.cazaar.com/ta/econ113/interpreting-beta
9. Kim, B. (2015, Sep 21). Understanding diagnostic plots for linear regression analysis. Retrieved from https://data.library.virginia.edu/diagnostic-plots/
10. Lim, K. (2017, Feb 14). Singapore visitor arrivals hit record propelled by Chinese, Indonesian tourists. *Nikkei Asian Review*. Retrieved from https://asia.nikkei.com/Economy/Singapore-visitor-arrivals-hit-record-propelled-by-Chinese-Indonesian-tourists2
11. Mayer, T. & Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist Database. *CEPII Working Paper 2011-25*
12. Qualtrics. (n.d.). Interpreting residual plots to improve your regression. Retrieved from https://www.qualtrics.com/support/stats-iq/analyses/regression-guides/interpreting-residual-plots-improve-regression/

# 6 References


12. Sars in Singapore: Timeline. (2013, Mar 16). *The Straits Times*. Retrieved from https://www.straitstimes.com/singapore/sars-in-singapore-timeline
13. Singapore Tourism Board. (n.d.). Overview. Retrieved from https://www.stb.gov.sg/content/stb/en/about-stb/overview.html
14. Soeriaatmadja, W. (2018, Nov 10). STB event in Jakarta attracts hundreds of millenials. *The Straits Times*. Retrieved from https://www.straitstimes.com/asia/se-asia/singapore-tourism-board-event-in-jakarta-attracts-hundreds-of-millennials
15. The World Bank. (2019). *GDP per capita (constant 2010 US$)* [Data file]. Retrieved from https://data.worldbank.org/indicator/NY.GDP.PCAP.KD
16. The World Bank. (2019). *Population, total* [Data file]. Retrieved from https://data.worldbank.org/indicator/SP.POP.TOTL?view=chart
17. Zhu, M. (2018, Mar 22). Travel & tourism forecast to attract nearly $21 bil in capital investments to Singapore in 2018. *The Edge Singapore*. Retrieved from https://www.theedgesingapore.com/travel-tourism-forecast-attract-nearly-21-bil-capital-investments-singapore-2018

# Thank You!

Ng Kai Hui | 18 Aug 2019 |
https://www.linkedin.com/in/ng-kai-hui/

# Instructions for use

**EDIT IN GOOGLE SLIDES**

Click on the button under the presentation preview that says "Use as Google Slides Theme".

You will get a copy of this document on your Google Drive and will be able to edit, add or delete slides.

You have to be signed in to your Google account.

**EDIT IN POWERPOINT®**

Click on the button under the presentation preview that says "Download as PowerPoint template". You will get a .pptx file that you can edit in PowerPoint.

Remember to download and install the fonts used in this presentation (you'll find the links to the font files needed in the Presentation design slide)

***More info on how to use this template at www.slidescarnival.com/help-use-presentation-template***

# Hello!

## *I am Jayden Smith*

I am here because I love to give presentations.

You can find me at @username

1

# Transition headline

Let’s start with the first set of slides

*Quotations are commonly printed as a means of inspiration and to invoke philosophical thoughts from the reader.*

## This is a slide title

- Here you have a list of items
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content, but won’t do both.

# Big concept

Bring the attention of your audience over a key concept using icons or illustrations

## You can also split your content

### White

Is the color of milk and fresh snow, the color produced by the combination of all the colors of the visible spectrum.

### Black

Is the color of coal, ebony, and of outer space. It is the darkest color, the result of the absence of or complete absorption of light.

# In two or three columns

**Yellow**

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

**Blue**

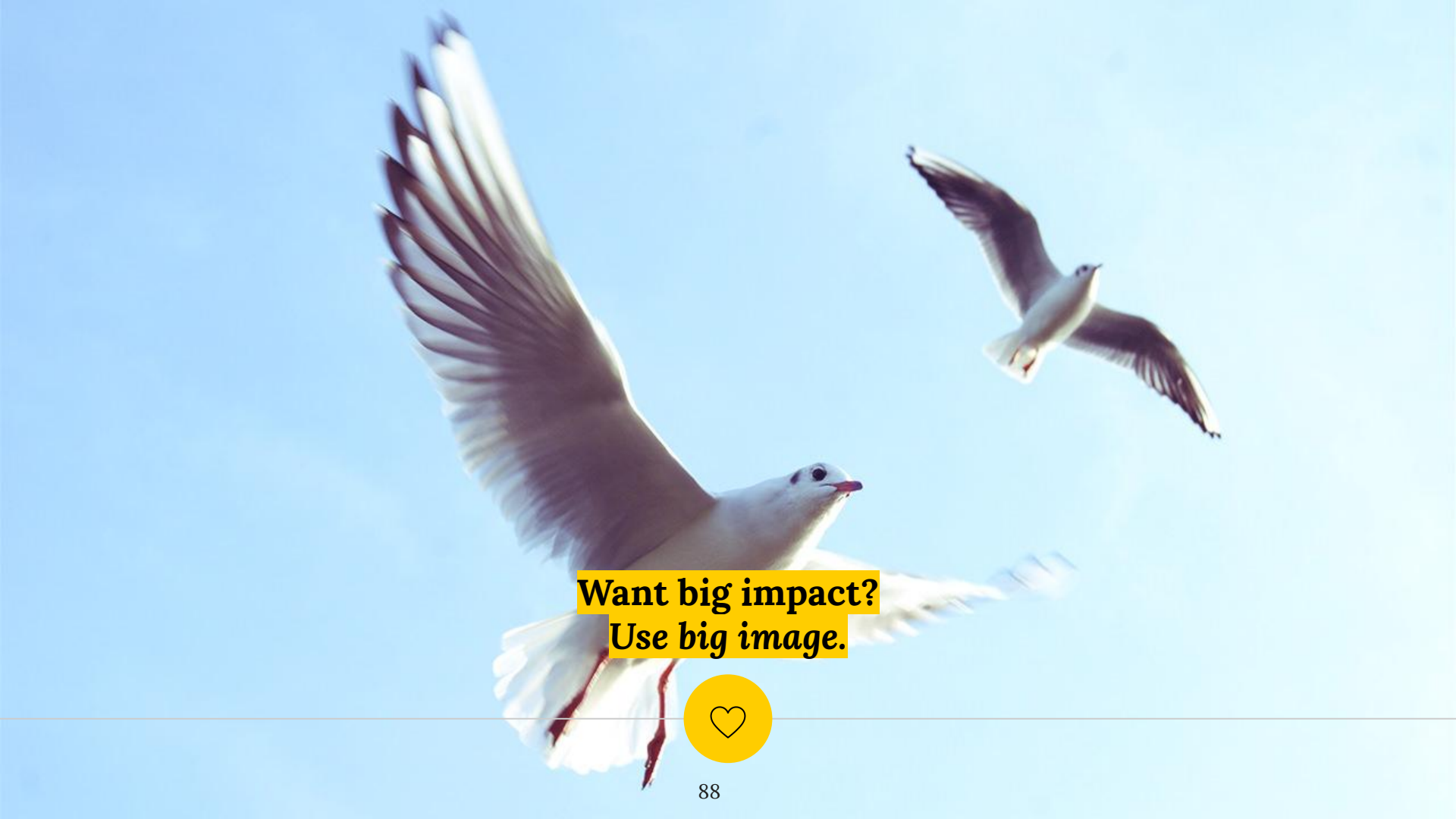
Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

**Red**

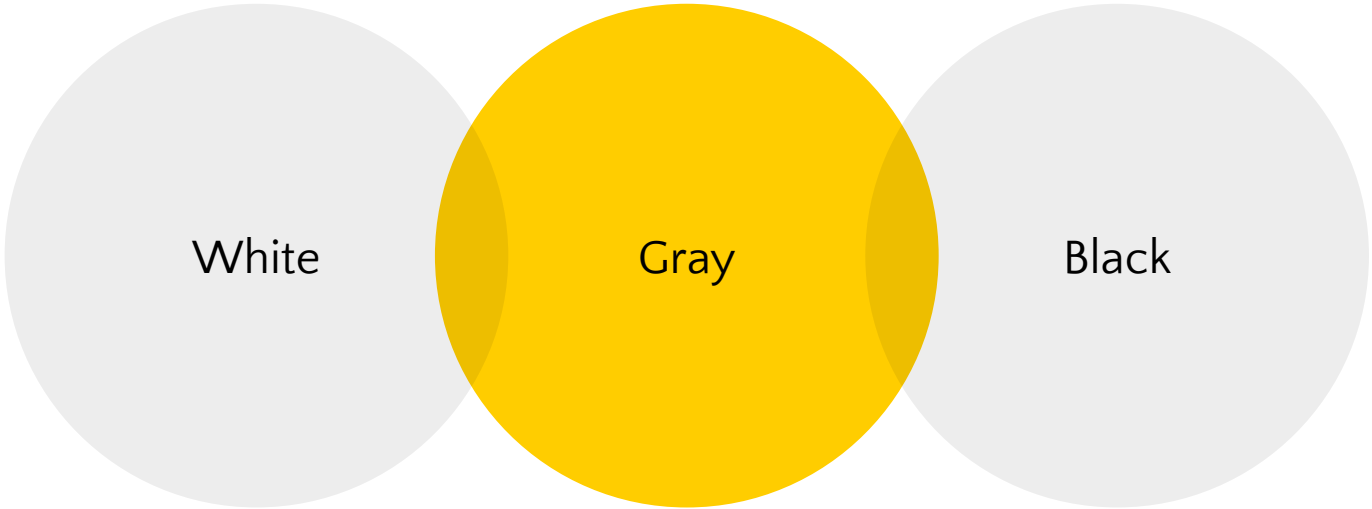
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

## A picture is worth a thousand words

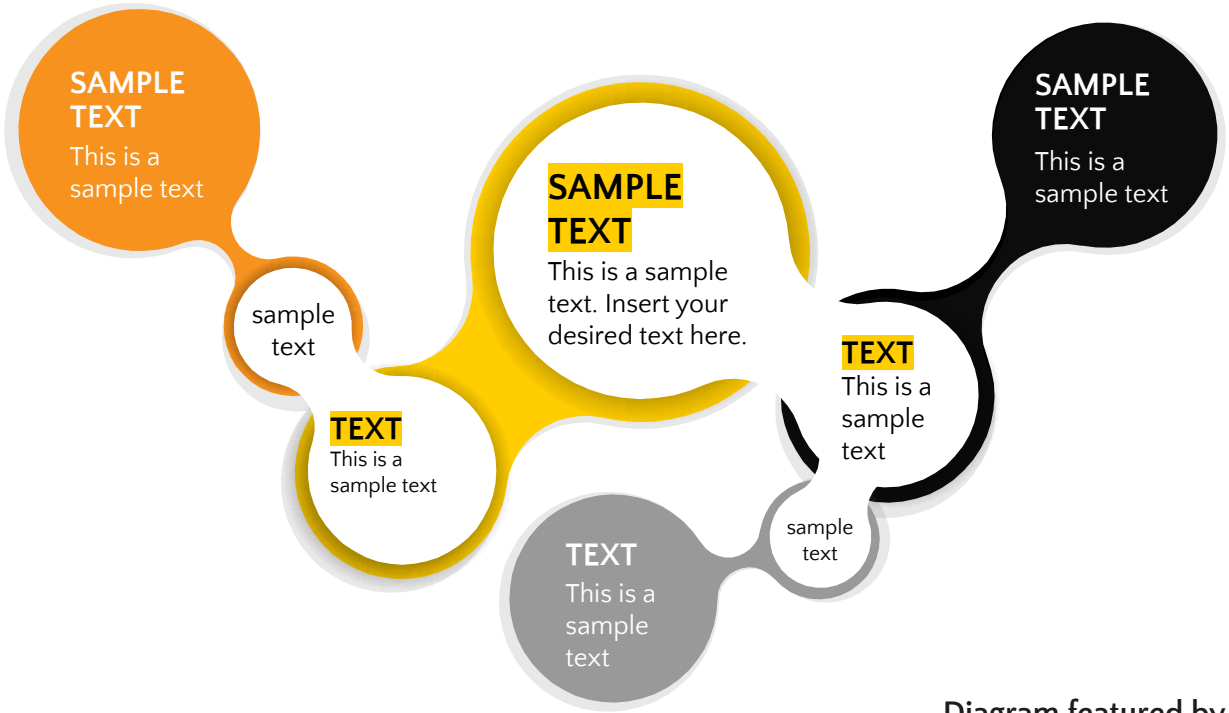
A complex idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.


Want big impact?
Use big image.

# Use charts to explain your ideas

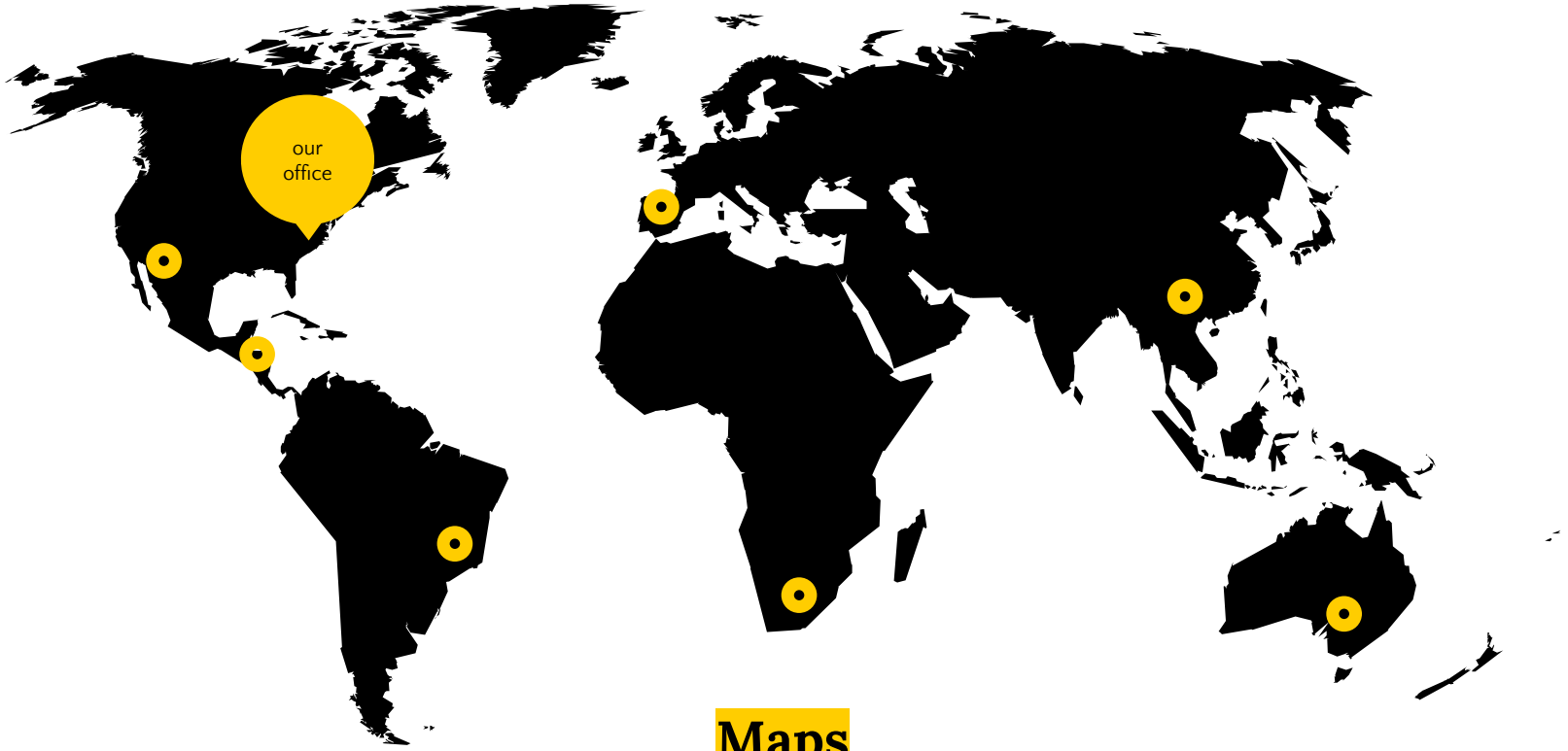


# Or use diagrams to explain complex ideas



# And tables to compare data

| | A | B | C |
|---|---|---|---|
| Yellow | 10 | 20 | 7 |
| Blue | 30 | 15 | 10 |
| Orange | 5 | 24 | 16 |




# Maps

# 89,526,124

Whoa! That’s a big number, aren’t you proud?

# 89,526,124$

That’s a lot of money

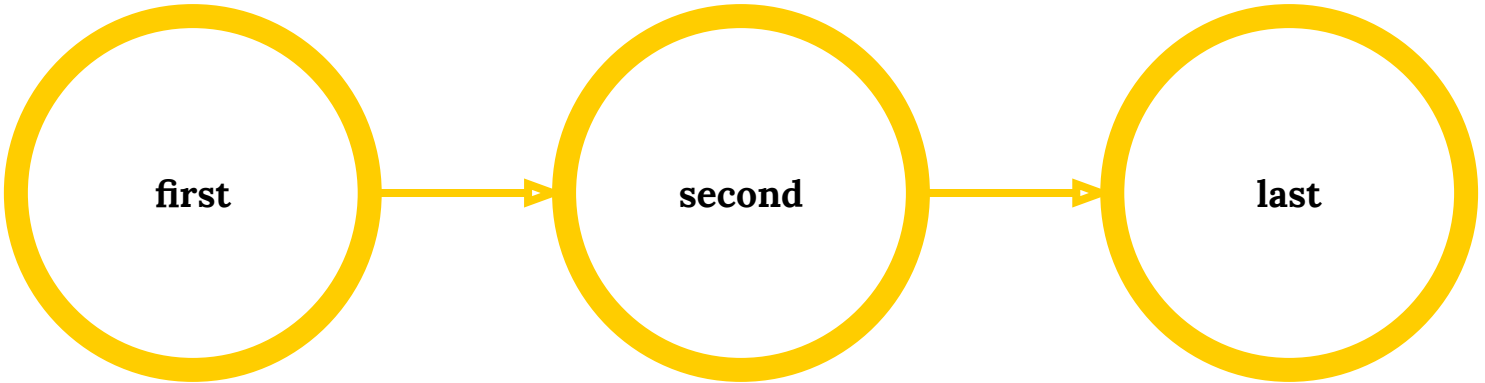
# 185,244 users

And a lot of users

# 100%

Total success!

# Our process is easy



# Let’s review some concepts

**Yellow**

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

**Blue**

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

**Red**

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

**Yellow**

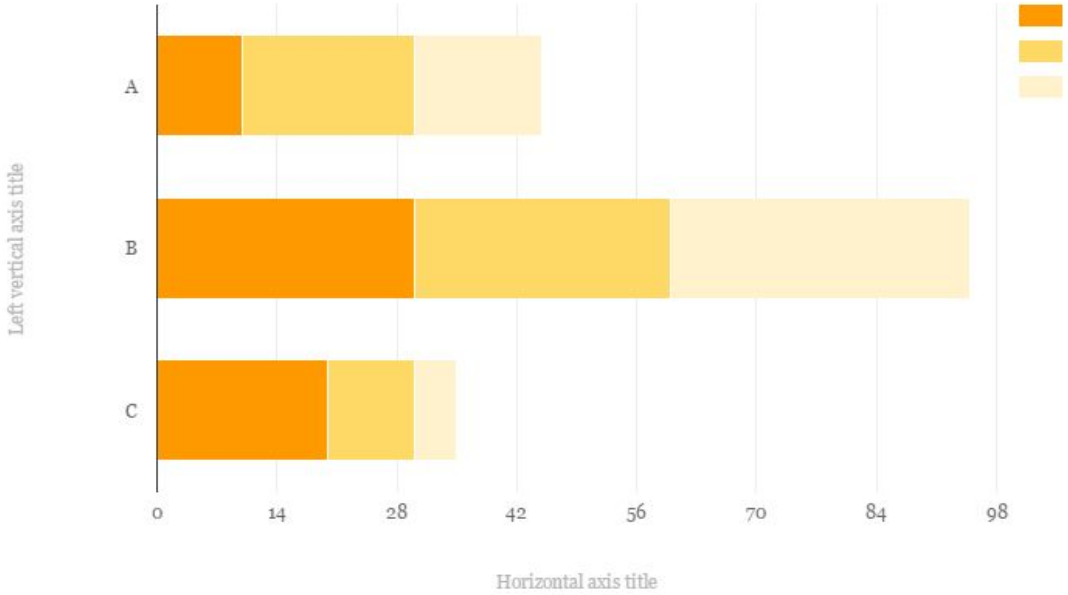
Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

**Blue**

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

**Red**

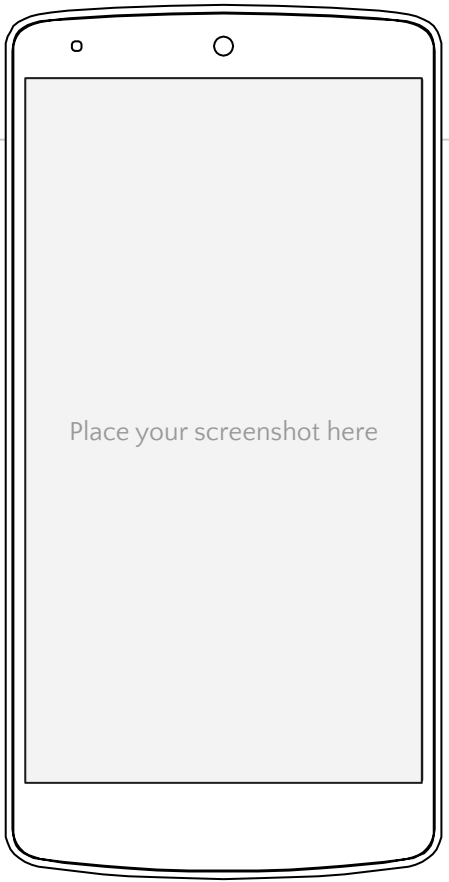
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.




*You can copy&paste graphs from Google Sheets*

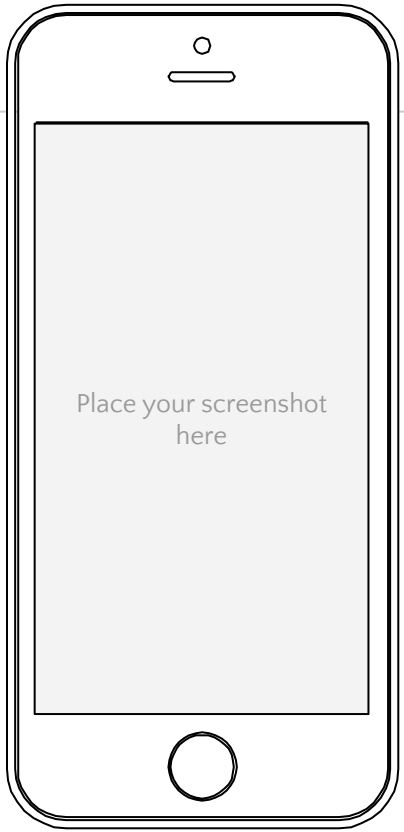
# Android project

Show and explain your web, app or software projects using these gadget templates.



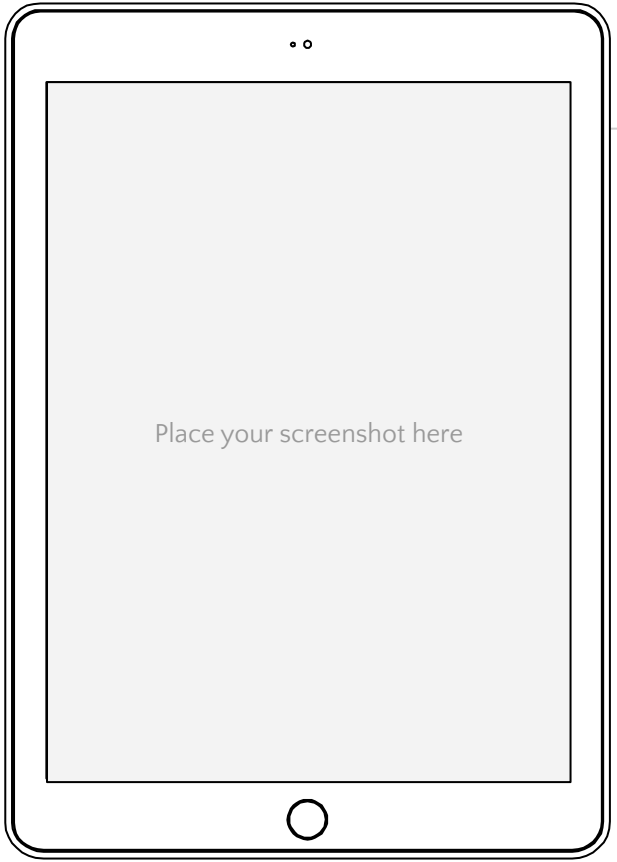
## iPhone project

Show and explain your web, app or software projects using these gadget templates.



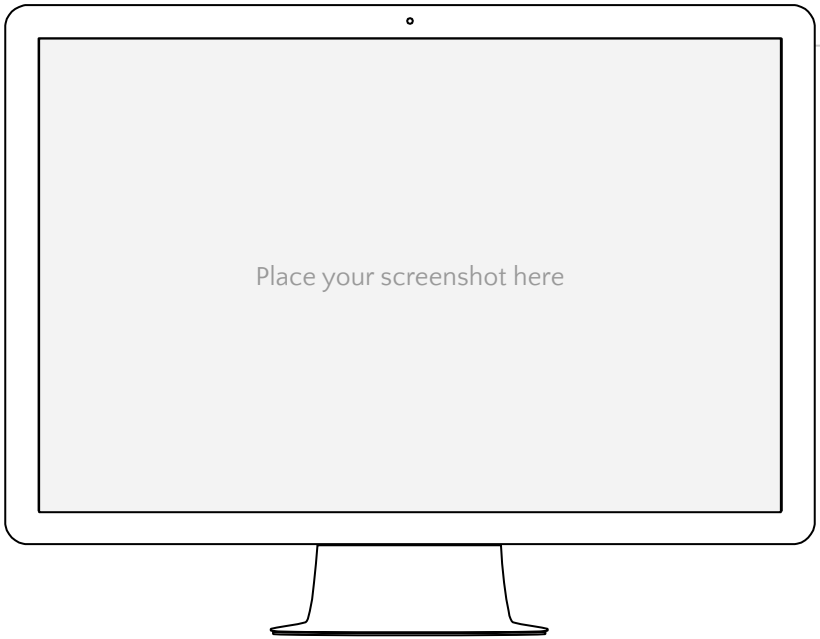
## Tablet project

Show and explain your web, app or software projects using these gadget templates.



## Desktop project

Show and explain your web, app or software projects using these gadget templates.



# Thanks!

## *Any questions ?*

You can find me at

- @username
- user@mail.me

# Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash

# Presentation design

This presentation uses the following typographies and colors:

- Titles: **Lora**
- Body copy: **Quattrocento Sans**

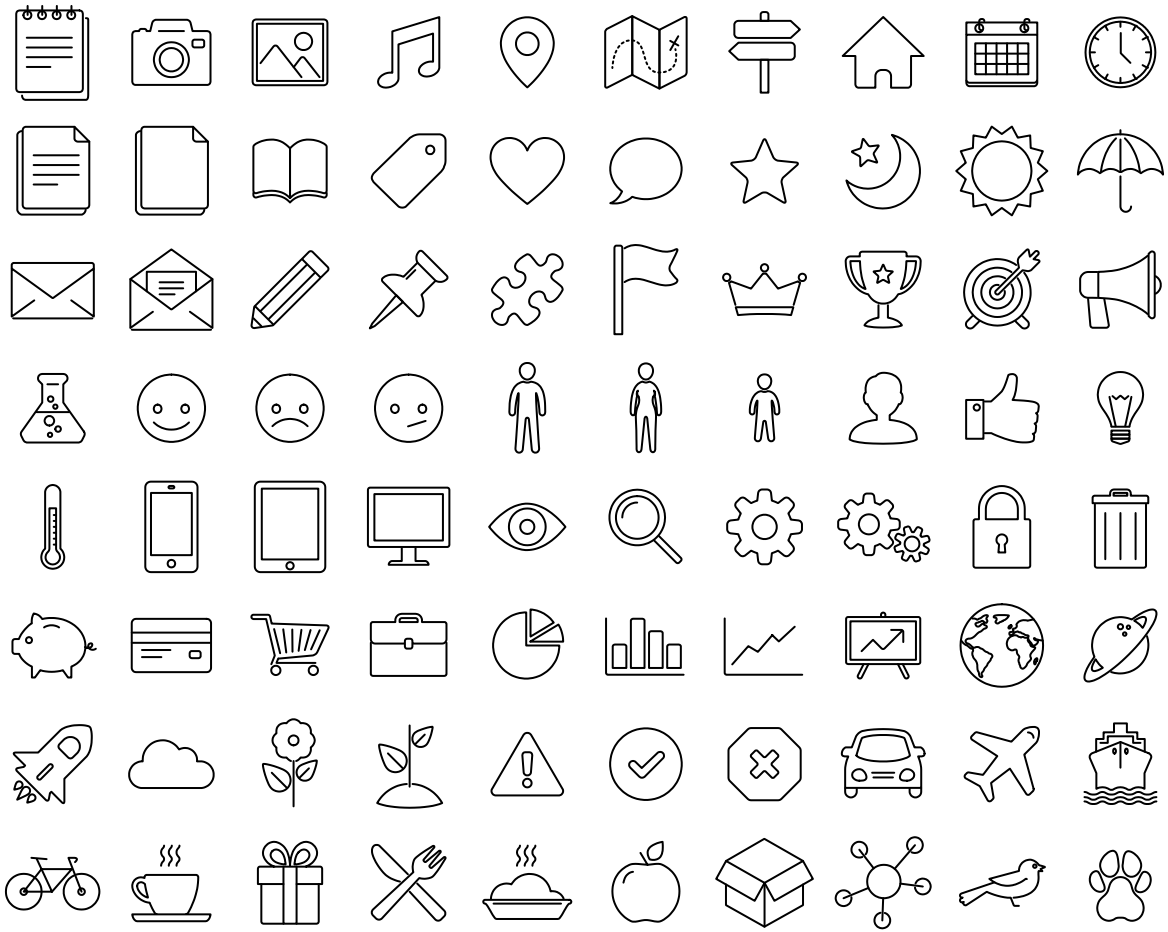
You can download the fonts on these pages:

https://www.fontsquirrel.com/fonts/lora

https://www.fontsquirrel.com/fonts/quattrocento-sans

Yellow #ffcd00 | Black **#000000** | Grey #cccccc

*You don't need to keep this slide in your presentation. It's only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®*



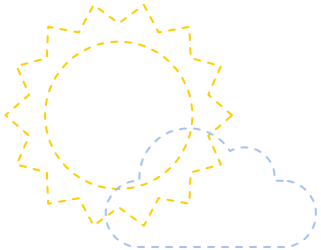
**SlidesCarnival icons are editable shapes.**

This means that you can:

- Resize them without losing quality.
- Change line color, width and style.

Isn't that nice? :)

Examples:



**Now you can use any emoji as an icon!**
And of course it resizes without losing quality and you can change the color.

How? Follow Google instructions
https://twitter.com/googledocs/status/730087240156643328

and many more...