Microsoft Sanddance

工程師的第一堂資料課
之毫無美感也可以玩視覺化

Mia Chang

tw MVC

About me

- Data Analyst / Data Consultant
  - Data Visualization
  - Algorithm Research
  - Machine Learning / Deep Learning
  - Co-Founder Azure Taiwan Community



Topics

工程師的第一堂資料課之 毫無美感也可以玩視覺化

- 大家在工作時的角色是什麼？
  - 沒有程式背景的初心者
  - 有程式背景想學資料分析
  - 平時已經在做資料分析
  - 其他

Topics

工程師的第一堂資料課之 毫無美感也可以玩視覺化

- 如何問問題(畫什麼)
  - DashBoard / Metrics
  - Market Insights
  - Market Segmentation
  - Business Strategy
- 如何畫圖(怎麼畫)
  - 看看別人怎麼畫
  - 要從哪裡開始學
- 解讀圖表(圖畫出來呈現的意義有哪些)
- 延伸閱讀
- Q&A

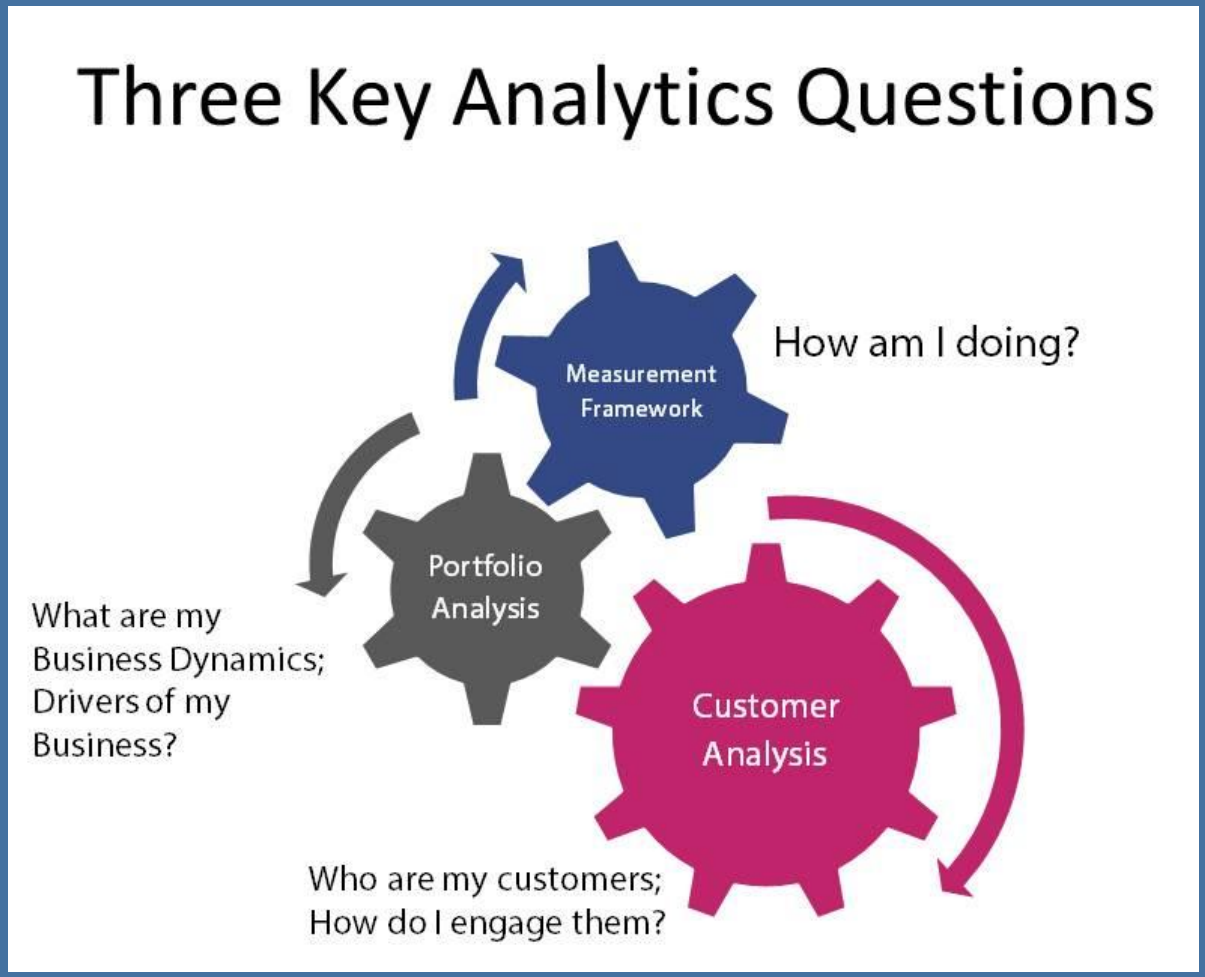


demo

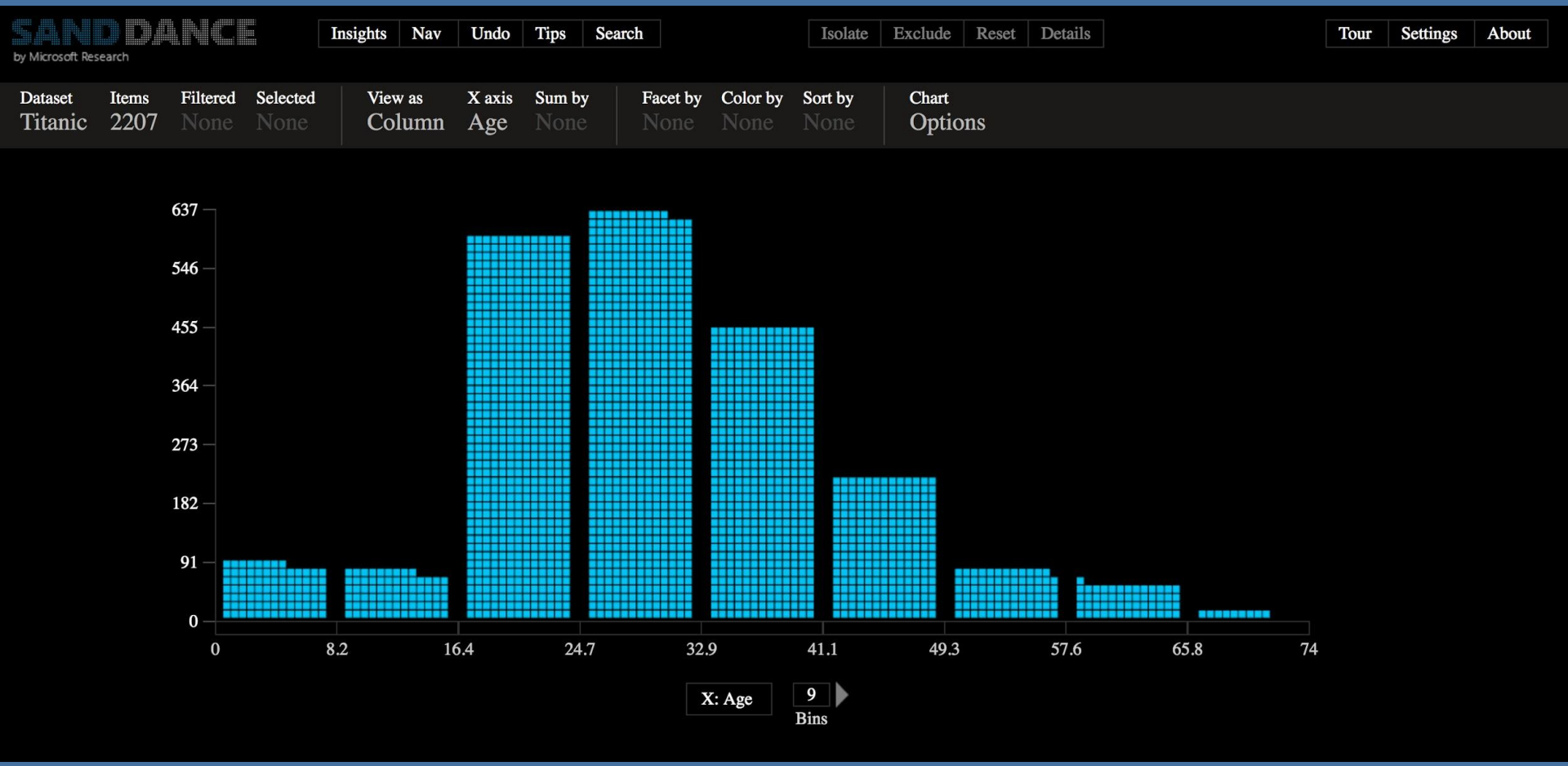
如何問問題

Data is everywhere

DashBoard / Metrics
Market Insights
Market Segmentation
Business Strategy

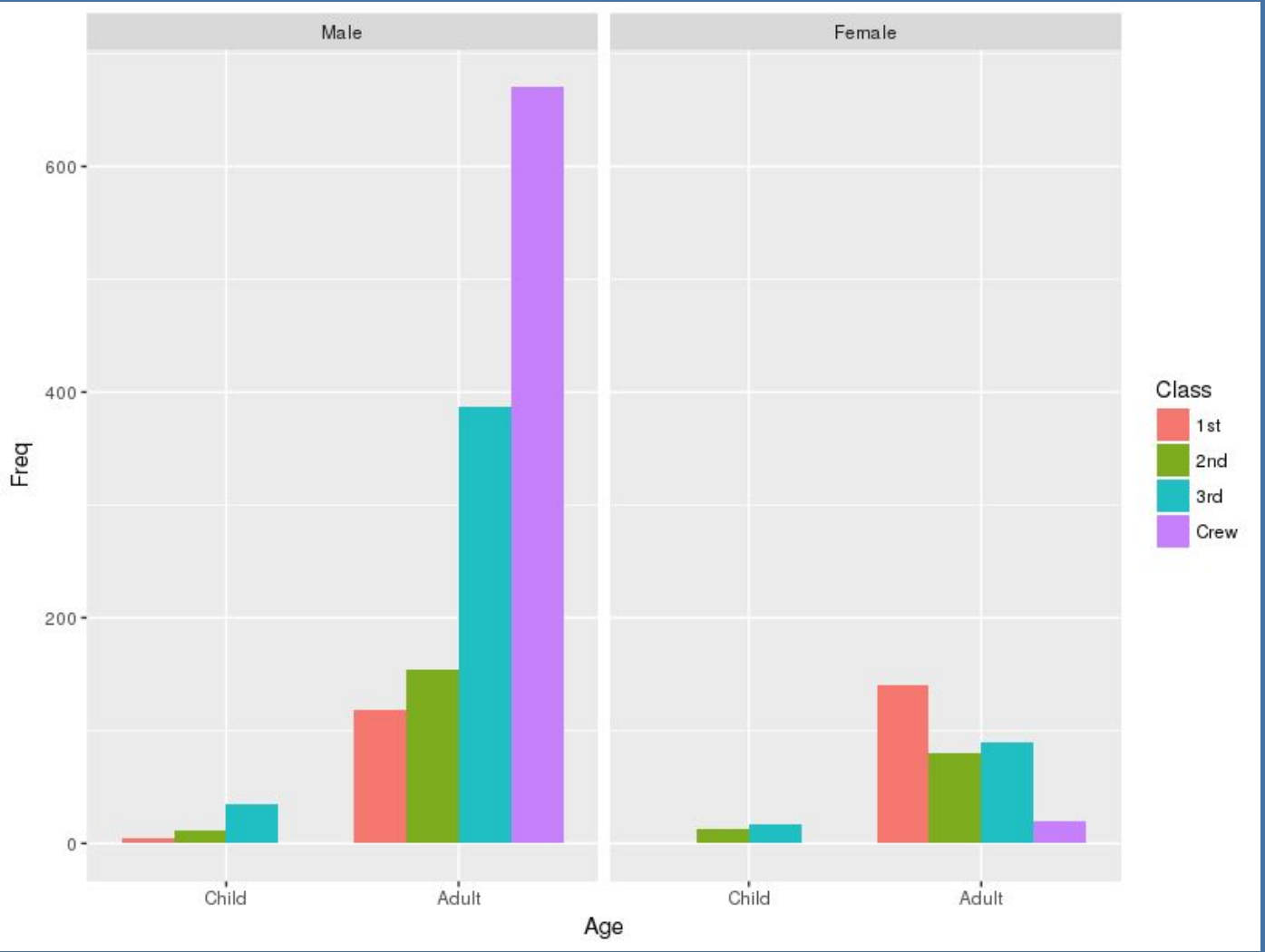


Data is everywhere



Data is everywhere

Role: 導演





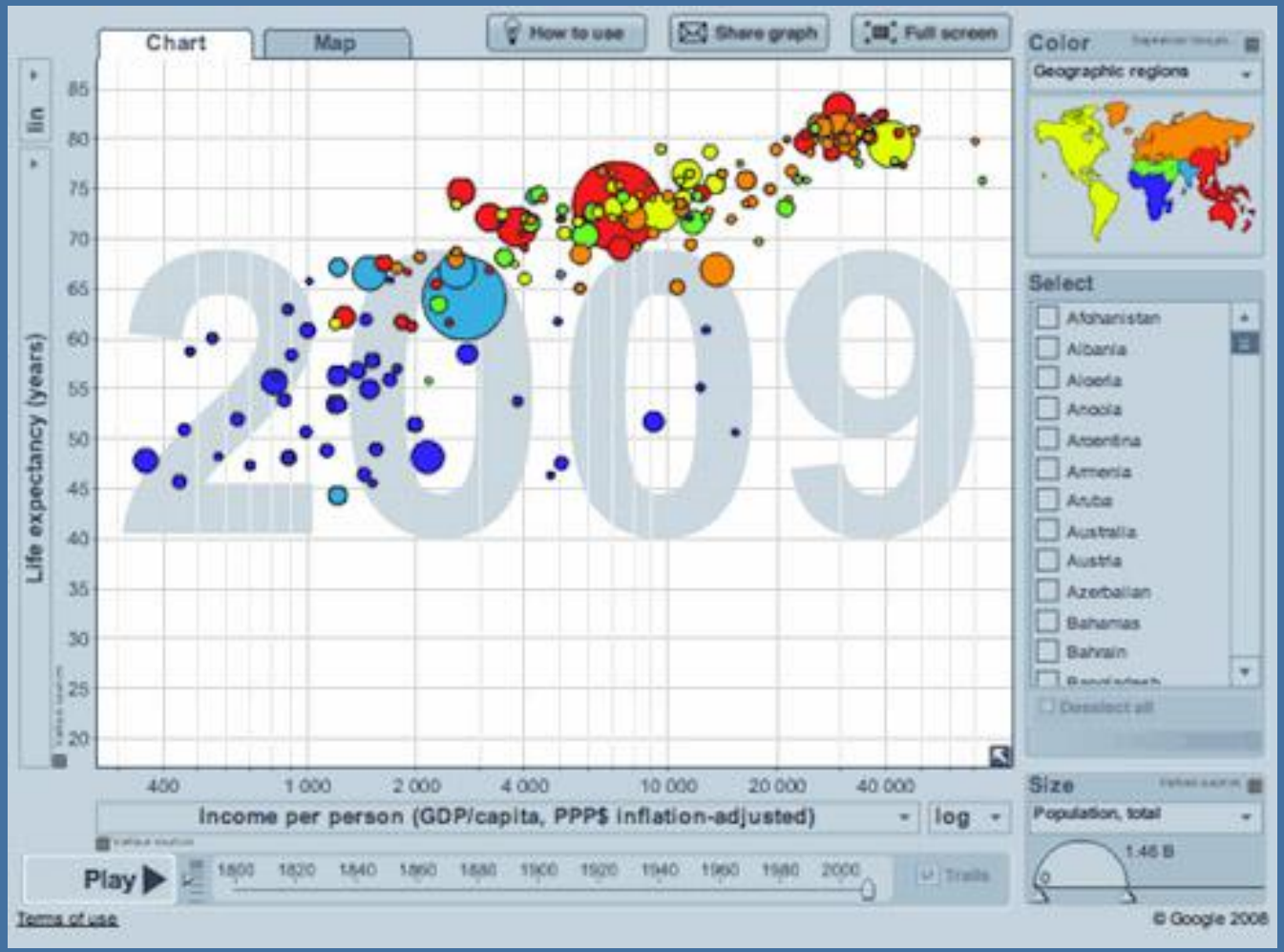
demo

如何畫圖

如何畫圖(怎麼畫)

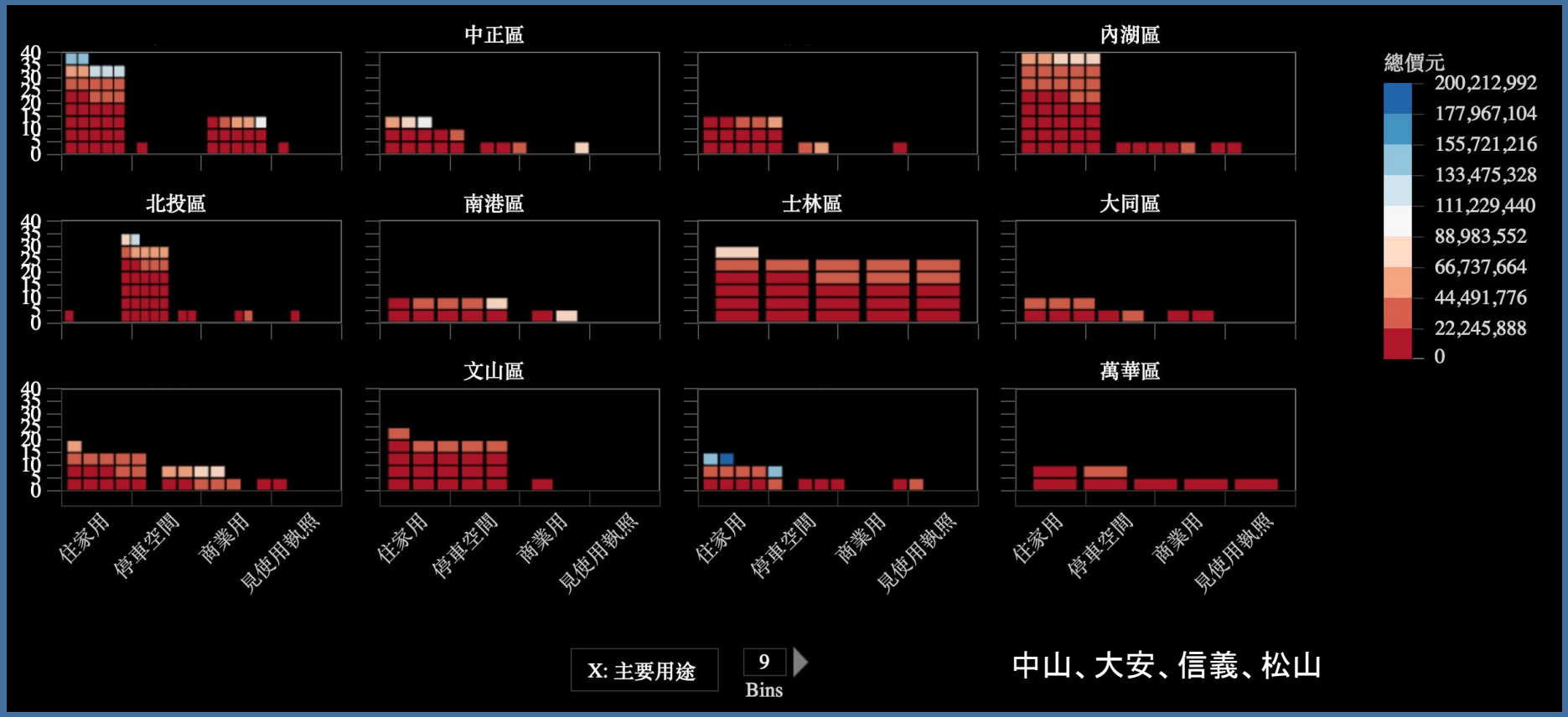
- 看看Demo Case, 看看別人怎麼畫
- 要從哪裡開始學

看看別人怎麼畫


Chart
Map
How to use
Share graph
Full screen
Color
Geographic regions
Select
Afghanistan
Albania
Algeria
Angola
Argentina
Armenia
Aruba
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Deselect all
Size
Population, total
1.46 B
Life expectancy (years)
lin
85
80
75
70
65
60
55
50
45
40
35
30
25
20
2009
400
1 000
2 000
4 000
10 000
20 000
40 000
Income per person (GDP/capita, PPP$ inflation-adjusted)
log
Play
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
Trails
Terms of use
© Google 2008

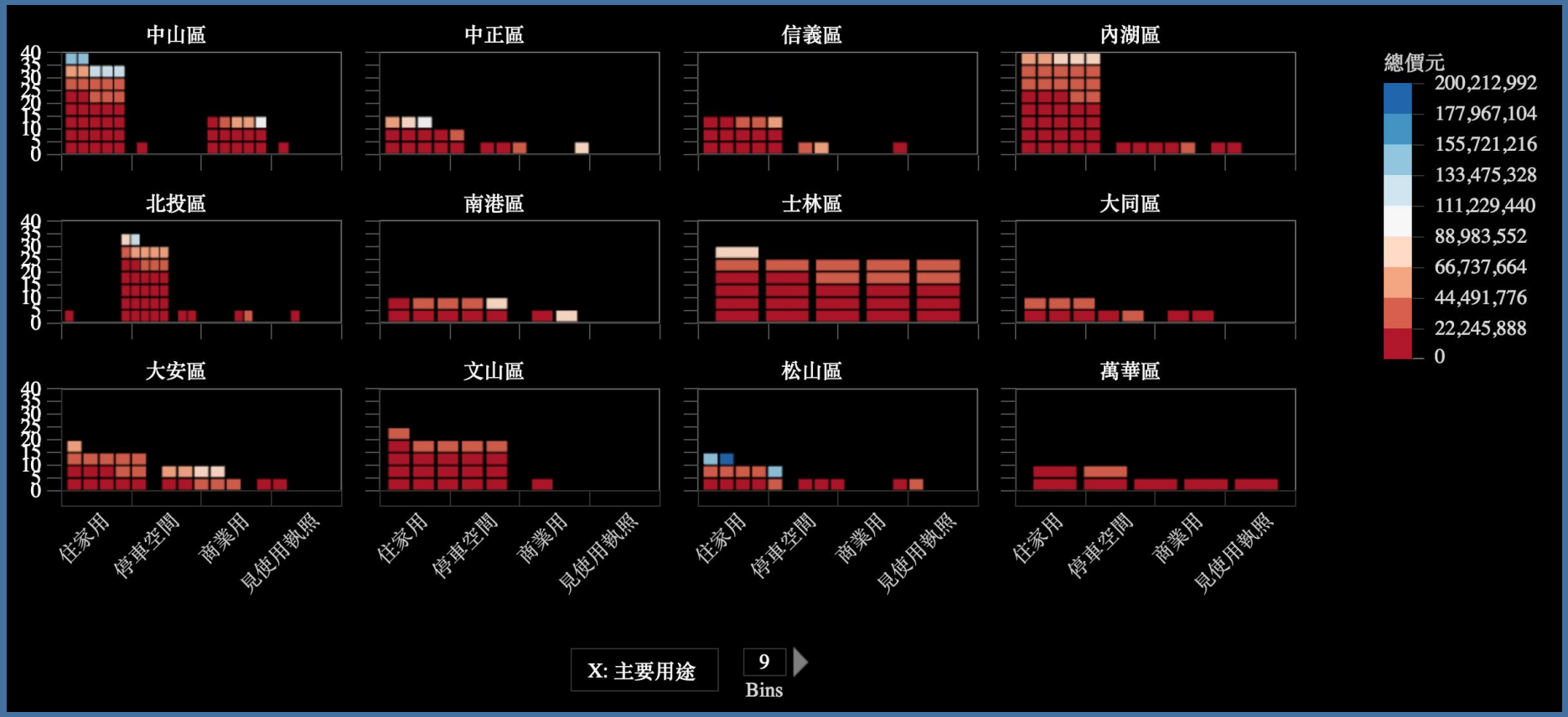
看看別人怎麼畫

Aesthetic Mapping



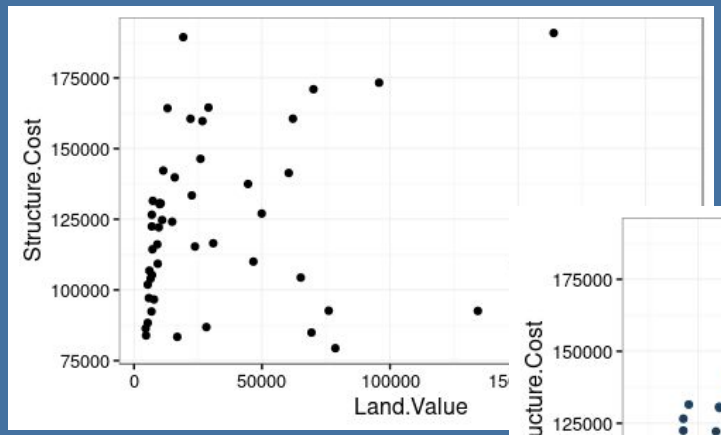
看看別人怎麼畫

Aesthetic Mapping

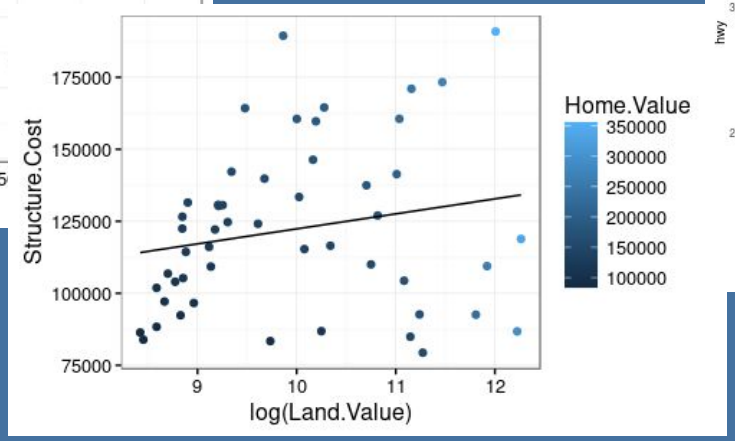


看看別人怎麼畫
Geometic Objects

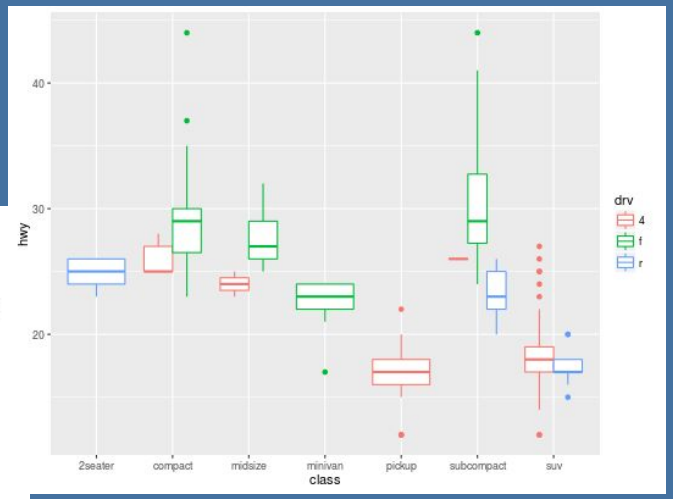
Points



Lines



Boxplot



看看別人怎麼畫
Data Format

Wide format (x)

IDNumbers	Name	Class1	Class2	Class3	Is Survived
001	Rose	1	0	0	1
002	Jack	0	1	0	0

Long format (v)

IDNumbers	Name	Class	Is Survived
001	Rose	1	1
002	Jack	2	0

要從哪裡開始學

One Variable

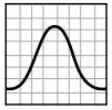
One Variable

Continuous

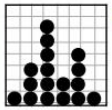
a <- ggplot(mpg, aes(hwy))



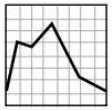
a + geom_area(stat = "bin")
x, y, alpha, color, fill, linetype, size
b + geom_area(aes(y = ..density..), stat = "bin")



a + geom_density(kernel = "gaussian"**)**
x, y, alpha, color, fill, linetype, size, weight
b + geom_density(aes(y = ..county..))



a + geom_dotplot()
x, y, alpha, color, fill



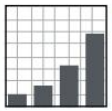
a + geom_freqpoly()
x, y, alpha, color, linetype, size
b + geom_freqpoly(aes(y = ..density..))



a + geom_histogram(binwidth = 5**)**
x, y, alpha, color, fill, linetype, size, weight
b + geom_histogram(aes(y = ..density..))

Discrete

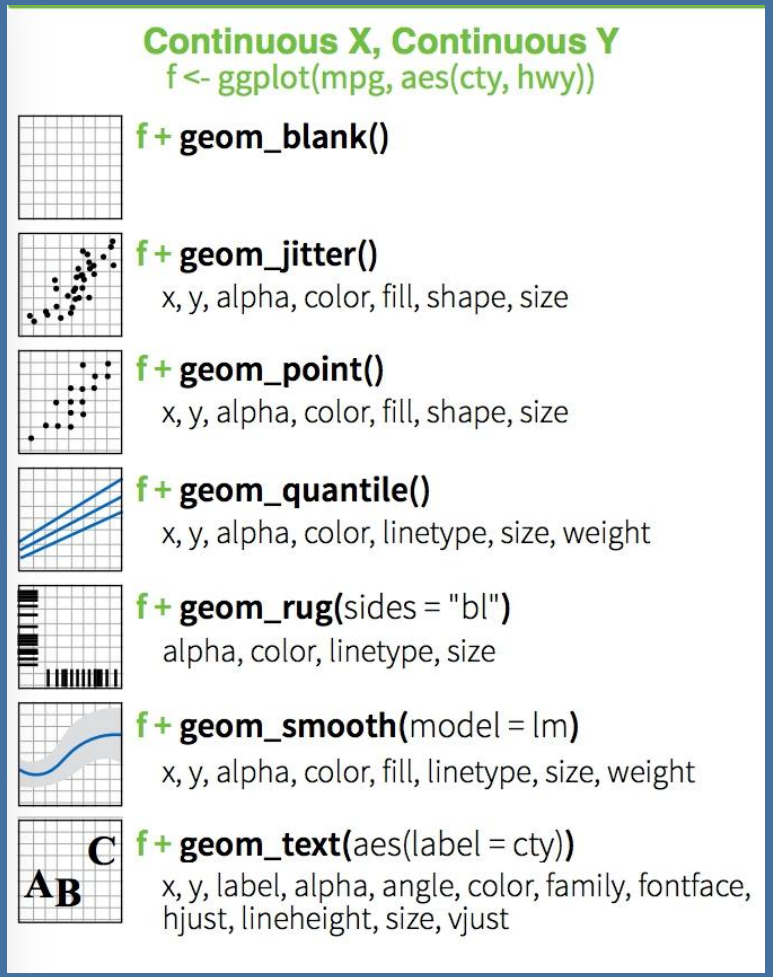
b <- ggplot(mpg, aes(fl))



b + geom_bar()
x, alpha, color, fill, linetype, size, weight

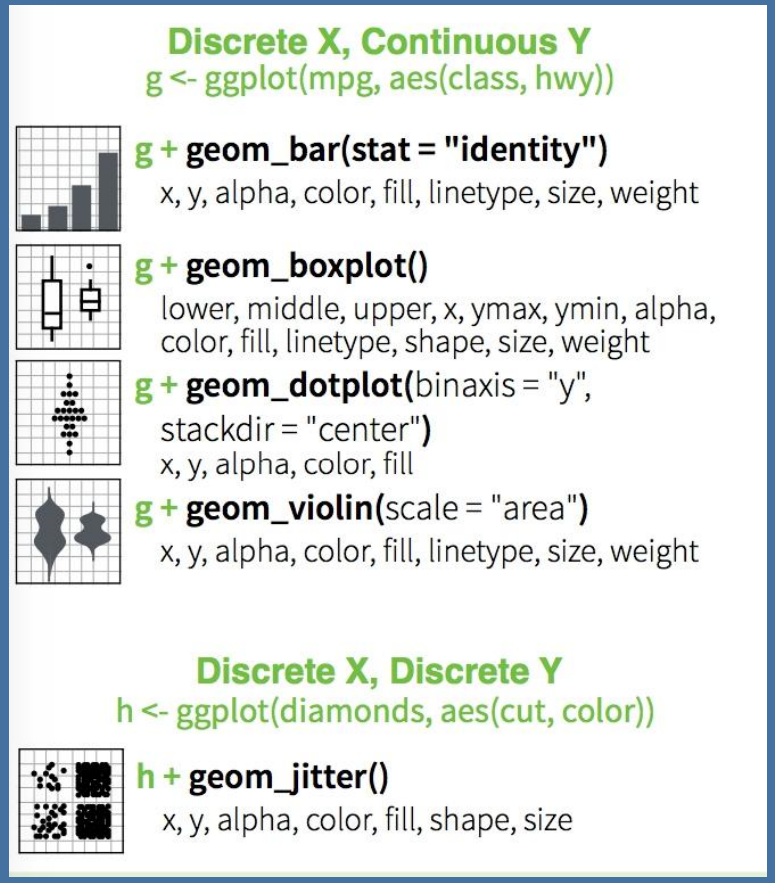
要從哪裡開始學

Two Variables



要從哪裡開始學

Two Variables



要從哪裡開始學

Two Variables

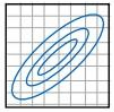
Continuous Bivariate Distribution

i <- ggplot(movies, aes(year, rating))



i + geom_bin2d(binwidth = c(5, 0.5)**)**

xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size, weight



i + geom_density2d()

x, y, alpha, colour, linetype, size



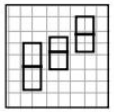
i + geom_hex()

x, y, alpha, colour, fill size

Visualizing error

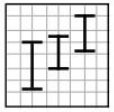
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)

k <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))



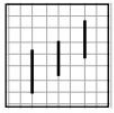
k + geom_crossbar(fatten = 2**)**

x, y, ymax, ymin, alpha, color, fill, linetype, size



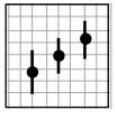
k + geom_errorbar()

x, ymax, ymin, alpha, color, linetype, size, width (also **geom_errorbarh()**)



k + geom_linerange()

x, ymin, ymax, alpha, color, linetype, size

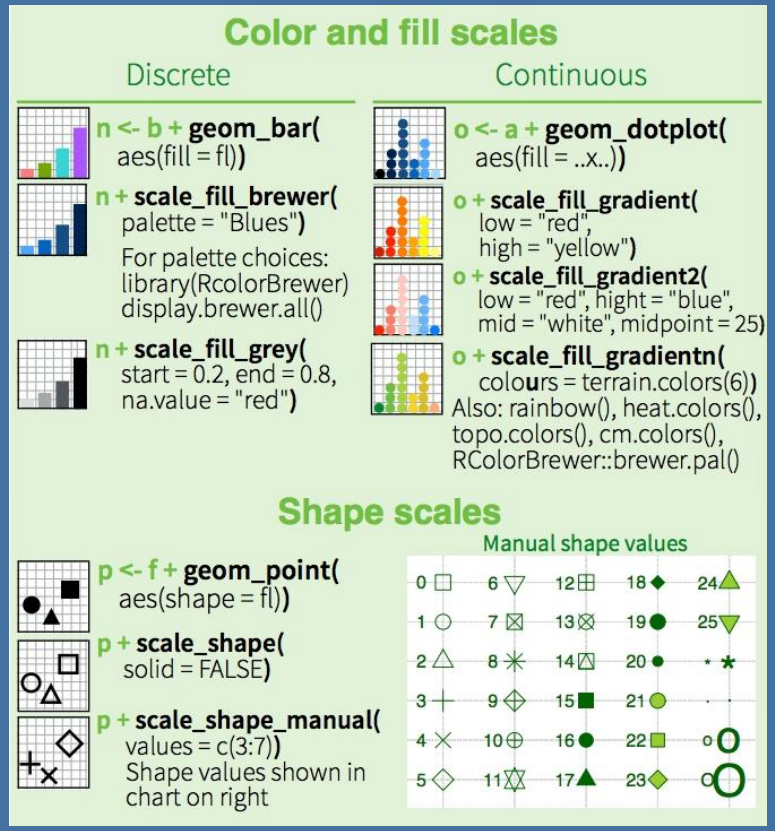


k + geom_pointrange()

x, y, ymin, ymax, alpha, color, fill, linetype, shape, size

要從哪裡開始學

Color / Shape





demo

解讀圖表

Answer your question

- 列出你的問題和Aesthetic Mapping之間的關聯
- 如何評估
  - Explain the data axes
  - Highlight subsets of data
  - Dig deeper to unwrap data
  - Place labels close to data points
  - Answer the “Why?” questions
  - Complement data with energetic delivery





demo

延伸閱讀

Read More...(Pic)

- 資料篩選
  - Logistic Regression for Rare Events
- 配色
- 分析模型（ Machine Learning / Deep Learning ）
  - 參數選擇 (Feature Engineering)

Read More...(Pic)

- TED
- https://www.ted.com/topics/visualizations

- Facebook Page
- https://www.facebook.com/Teradata/
- https://www.facebook.com/GraphicsSpeak/
- https://www.facebook.com/kaggle/
- https://www.facebook.com/visualnews/

- Reference
- 如何問問題
  - BerkeleyX: BUSADM466.2x Marketing Analytics: Price and Promotion Analytics
- 如何對應資料特徵 值
  - https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
- six-simple-techniques-for-presenting-data
  - http://sixminutes.dlugan.com/six-simple-techniques-for-presenting-data-hans-rosling-ted-2006/



demo

Q&A

Contact

- Facebook : facebook.com/mia5419
- Email : mia5419@gmail.com
- 幫徵才。

史上最強的注音輸入法



千呼萬喚

MAC 平台上，最聰明的注音輸入法
現在立即前往購買 ！

樂在Coding 愛上MVC

完全支援

請放心在 Windows 8 上使用 新自然輸入法 10 系統升級免煩惱，平板模式也沒問題！

Blog 是記錄知識的最佳平台

DotBlog

IT 技術人的部落格平台

https://dotblog.com.tw



感謝 Jetbrains 贊助贈品



ReSharper

https://www.jetbrains.com/resharper/

感謝 OzCode 贊助贈品



http://www.oz-code.com/

感謝 ALIVE 贊助贈品



https://comealive.io/

業界師資、實戰教學



http://skilltree.my

謝謝各位

http://mvc.tw

- 本投影片所包含的商標與文字皆屬原著作者所有。
- 本投影片使用的圖片皆從網路搜尋。
- 本著作係採用姓名標示 -非商業性-相同方式分享 3.0 台灣授權。閱讀本授權條款，請到 http://creativecommons.org/licenses/by-nc-sa/3.0/tw/，或寫信至Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.