Modern Data Science Methods for Educational Research

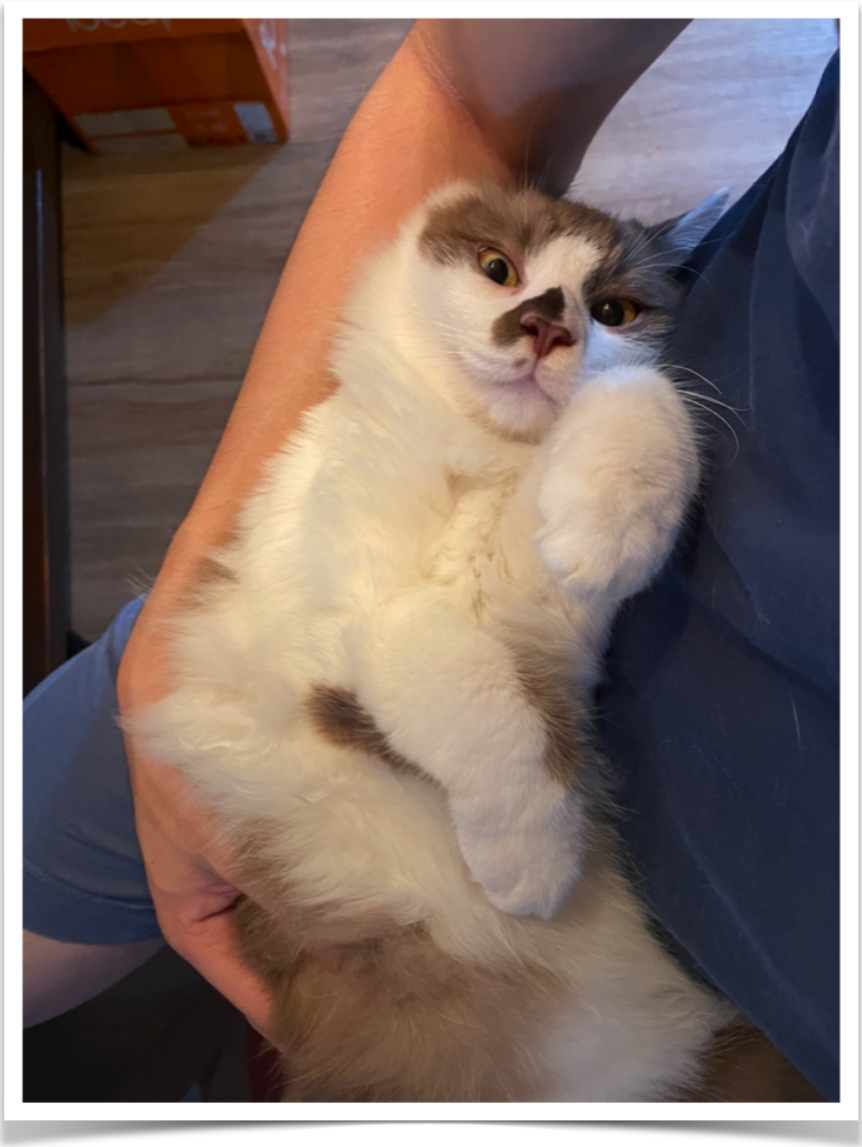
Creating Effective Data Visualization with R

Aesthetic Mapping and Scales

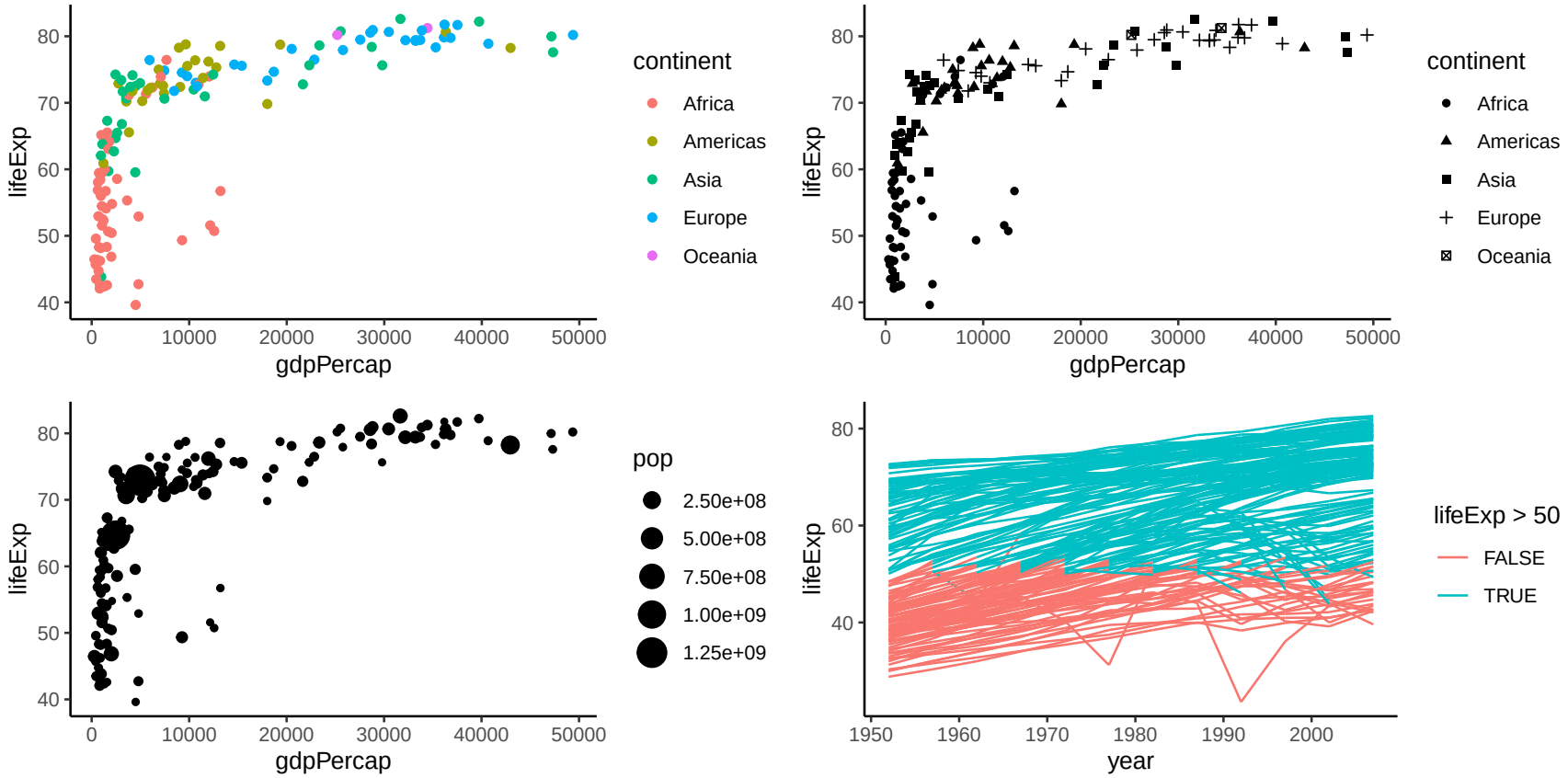
ผศ.ดร.สิวะโชติ ศรีสุทธิยากร

ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

February 18, 2023



Aesthetic mapping


lifeExp
gdpPercap
continent
Africa
Americas
Asia
Europe
Oceania
pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09
year
lifeExp > 50
FALSE
TRUE

Aesthetic mapping “Color”

กิจกรรม Aesthetic mapping: color

จากกิจกรรม Basic ggplot2

1. ขอให้ผู้เรียนทดลองกำหนด aesthetic mapping ของ color โดยจับคู่กับทวีปของประเทศ **ผลลัพธ์ที่ได้เป็นอย่างไร? ท่านคิดว่าสีที่กำหนดในทัศนภาพข้อมูลมีความเหมาะสมหรือไม่ เพราะเหตุใด?**
2. ลองจับคู่สี กับ จำนวนประชากรของประเทศ จากนั้นสังเกตผลลัพธ์ที่ได้ **มีความเหมือนหรือแตกต่าง กับการจับคู่กับทวีปของประเทศ อย่างไร เพราะเหตุใดจึงเป็นเช่นนั้น?**

กิจกรรม Aesthetic mapping: color

Syntax เดิม

```
gapminder %>% filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y = lifeExp))+
      geom_point()
```

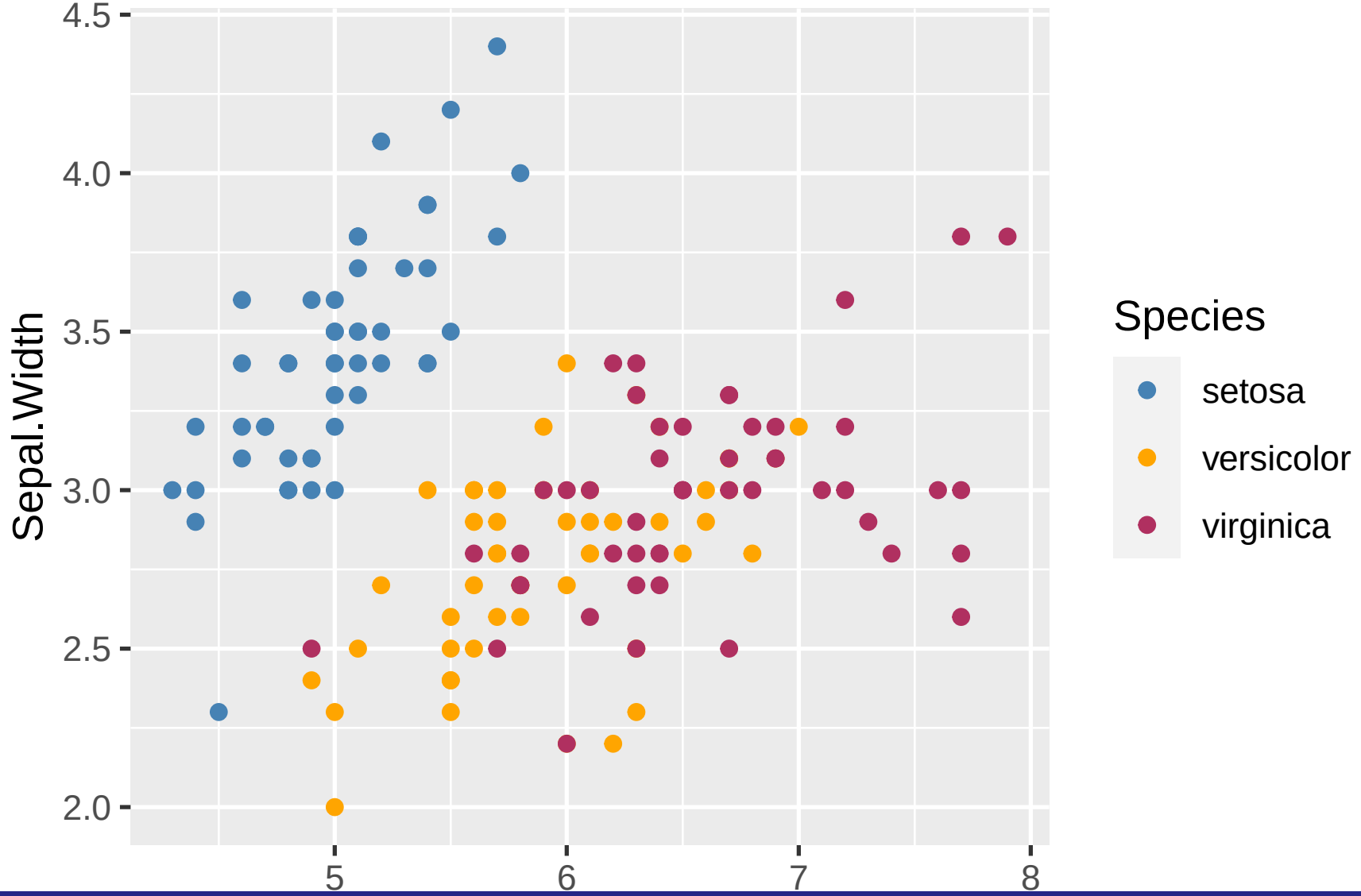
scale_color_XXX()

- scale_color_manual(name, values)
- scale_color_gradient(name, low, hight)
- scale_color_gradient2(name, low, mid, hight)

ตัวอย่างการใช้ฟังก์ชัน scale_color_XXX()

```
iris %>% ggplot(aes(x = Sepal.Length,
                    y = Sepal.Width,
                    col = Species))+
  geom_point()+
  scale_color_manual(values = c("steelblue",
                                "orange",
                                "maroon"))
```

ตัวอย่างการใช้ฟังก์ชัน `scale_color_XXX()`



แหล่งอ้างอิงสี

```
colors() %>% length()
```

```
[1] 657
```

```
colors() %>% head(12)
```

```
 [1] "white"         "aliceblue"     "antiquewhite"  "antic
 [5] "antiquewhite2" "antiquewhite3" "antiquewhite4" "aquar
 [9] "aquamarine1"   "aquamarine2"   "aquamarine3"   "aquar
```

แหล่งอ้างอิงสี

- https://colorhunt.co/
- https://coolors.co/
- https://www.canva.com/colors/color-palettes/
- https://color.adobe.com/create/color-wheel

กิจกรรม scale color

ขอให้ใช้ `<scale function>` ตัวใดก็ได้
กำหนดชุดของสีที่ใช้ในแผนภาพจากกิจกรรมที่ผ่านมา
โดยเลือกสีตามที่ผู้เรียนเห็นว่าเหมาะสมจากแหล่งอ้างอิงที่ให้ **ผลที่ได้เป็นอย่างไร**

Color is Matter?

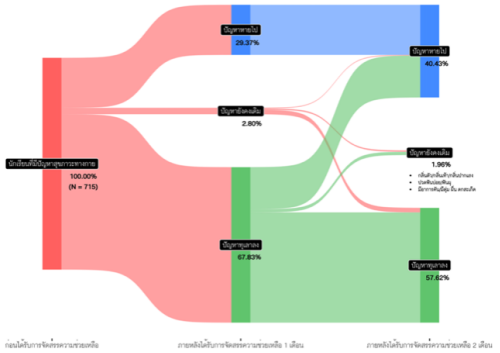
Color is Matter!!

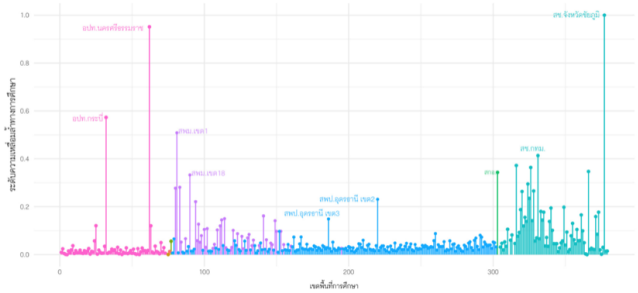
https://purple11.com/basics/hue-saturation-lightness/

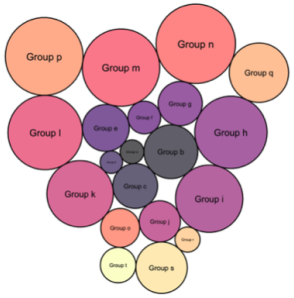
การใช้สีในทัศนภาพข้อมูล มีวัตถุประสงค์อยู่ 3 ลักษณะ

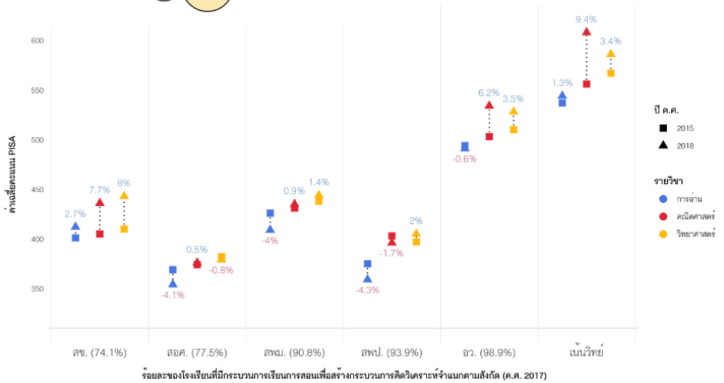
- Distinguish tools
- Represent data values
- Highlight

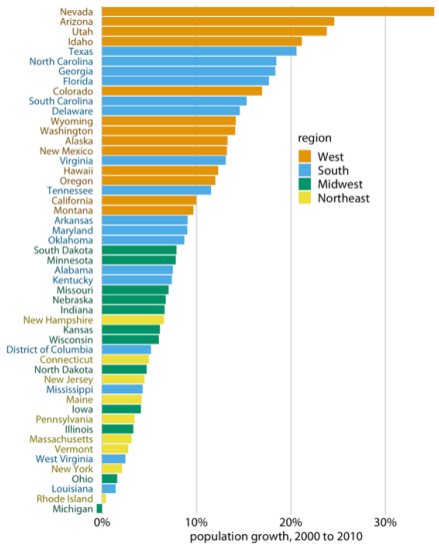
Color as Distinguish tools







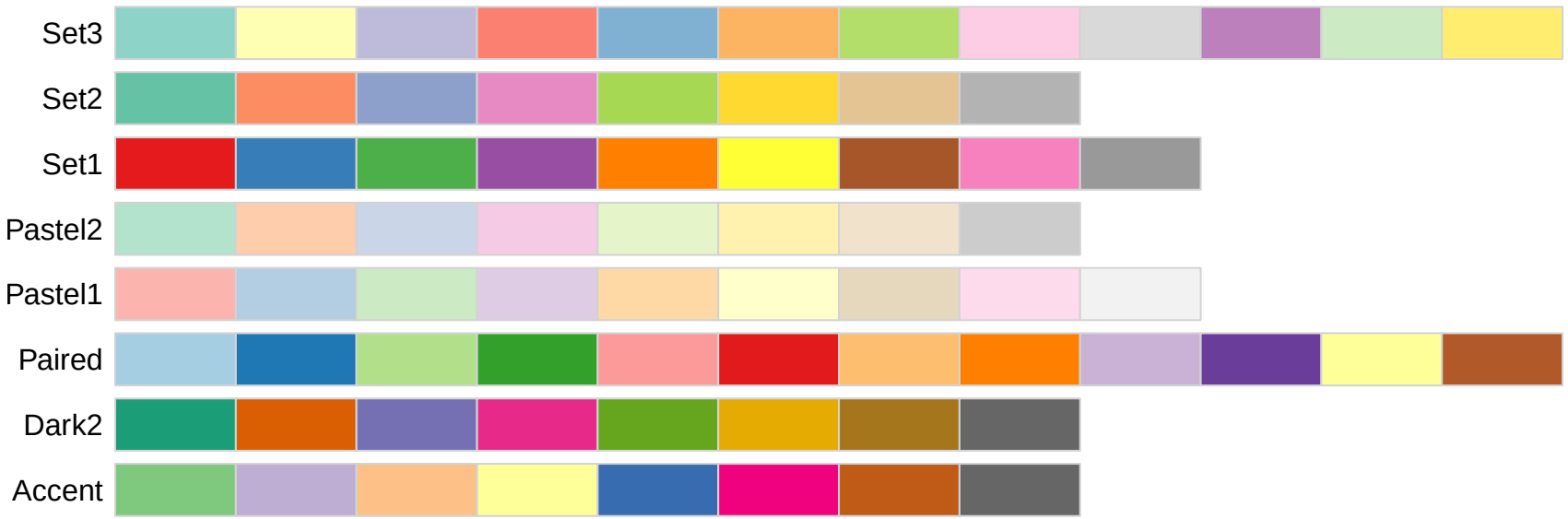




(Wilke, 2019)

Color as Distinguish tools: Qualitative palette

```r
library(RColorBrewer)
display.brewer.all(type = "qual")
```



Color to represent data values

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

FIGURE 4.15 Use color sparingly

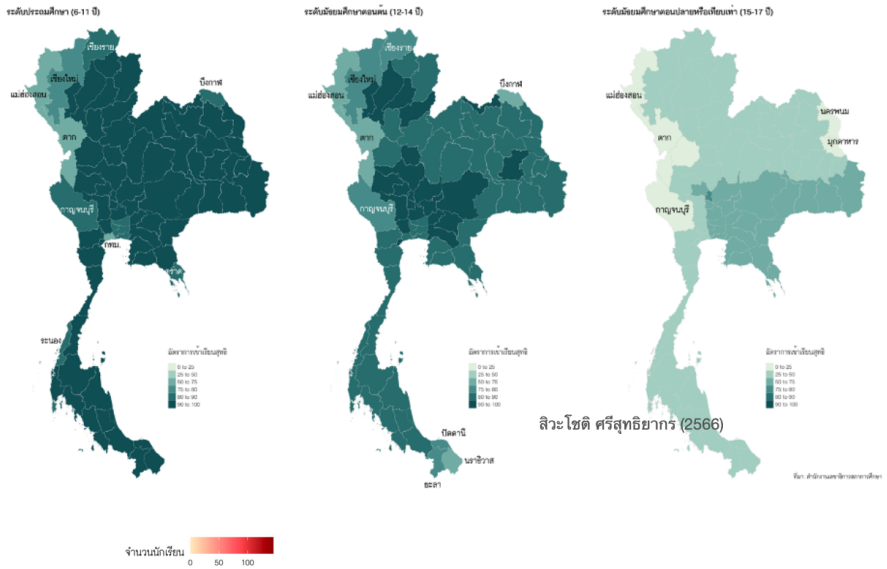
Top 5 drugs: country-level sales rank

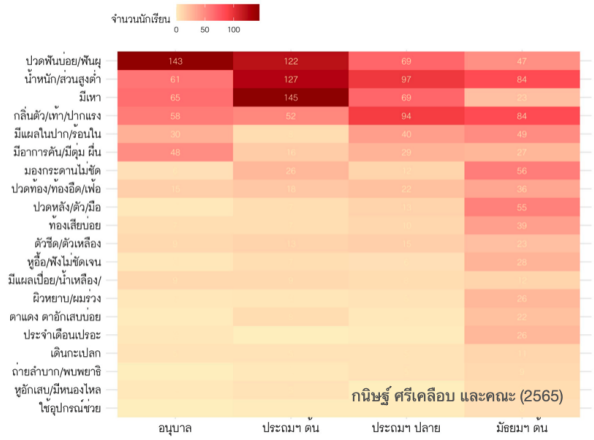
RANK: 1 | 2 | 3 | 4 | 5+

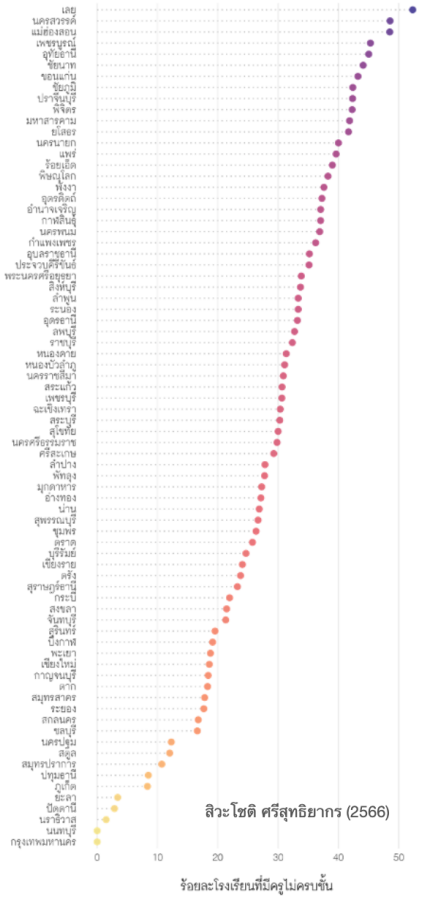
COUNTRY | DRUG	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5

(Knaflic, 2015)

Color to represent data values

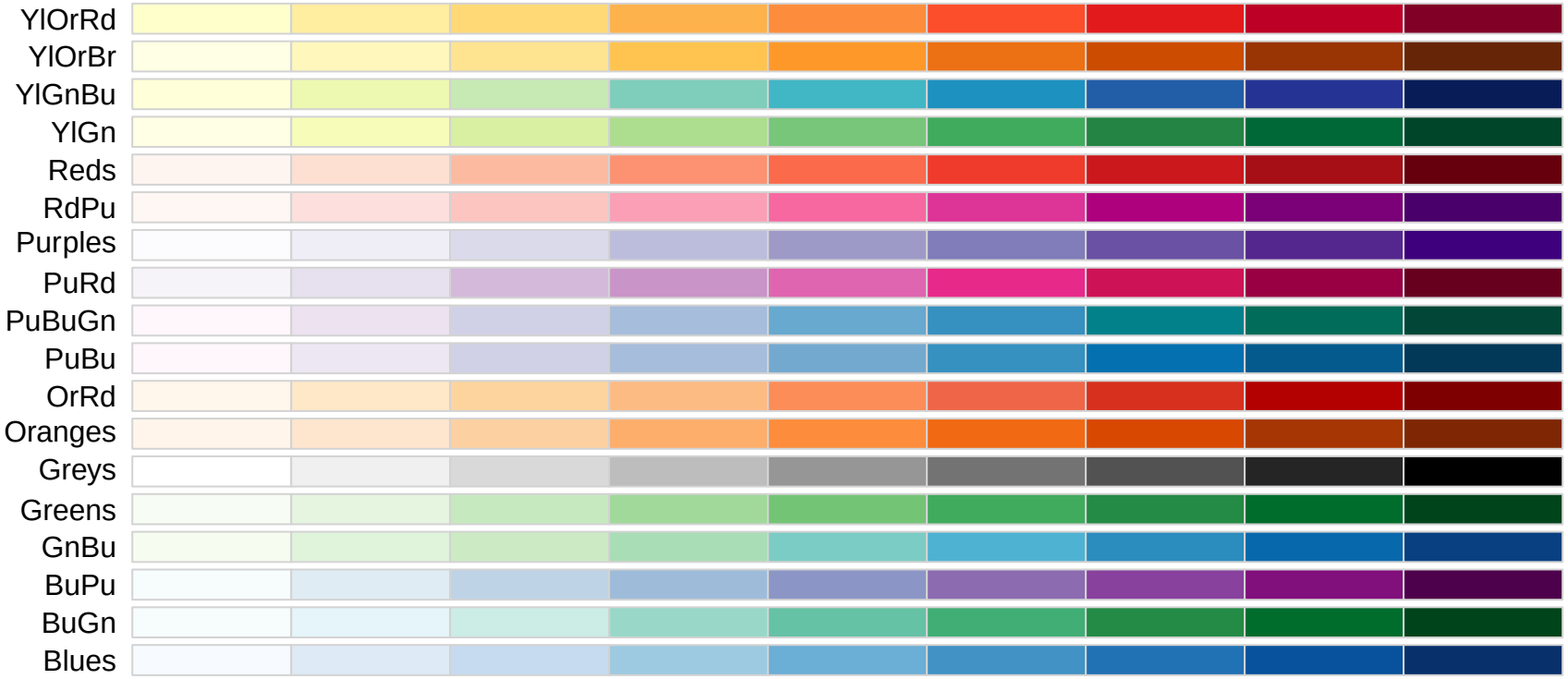




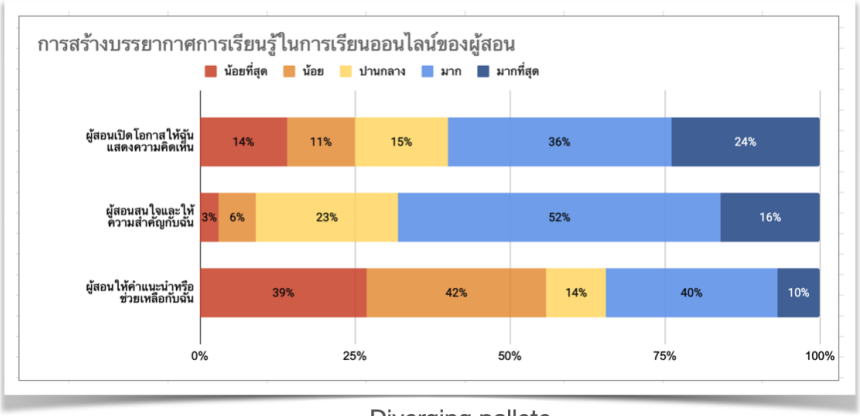


Color to represent data values: Sequential palette

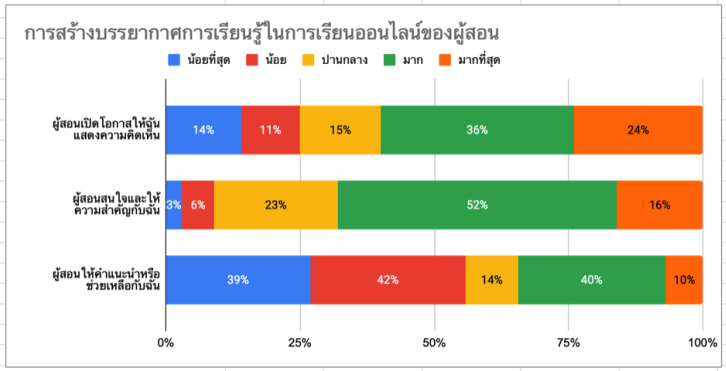
```
display.brewer.all(type = "seq")
```



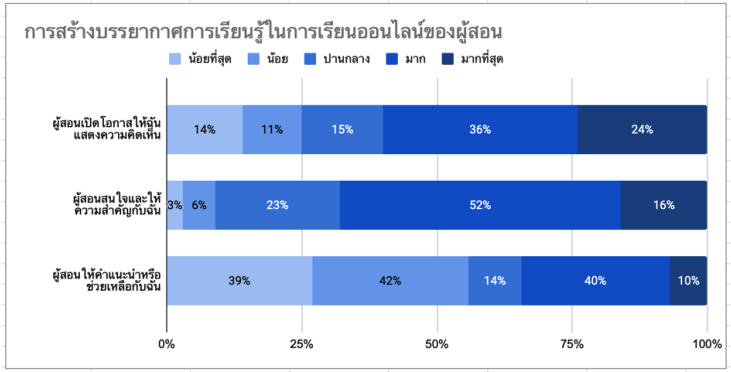
Color to represent data values: Diverging palette



Diverging pallete



Qualitative pallete

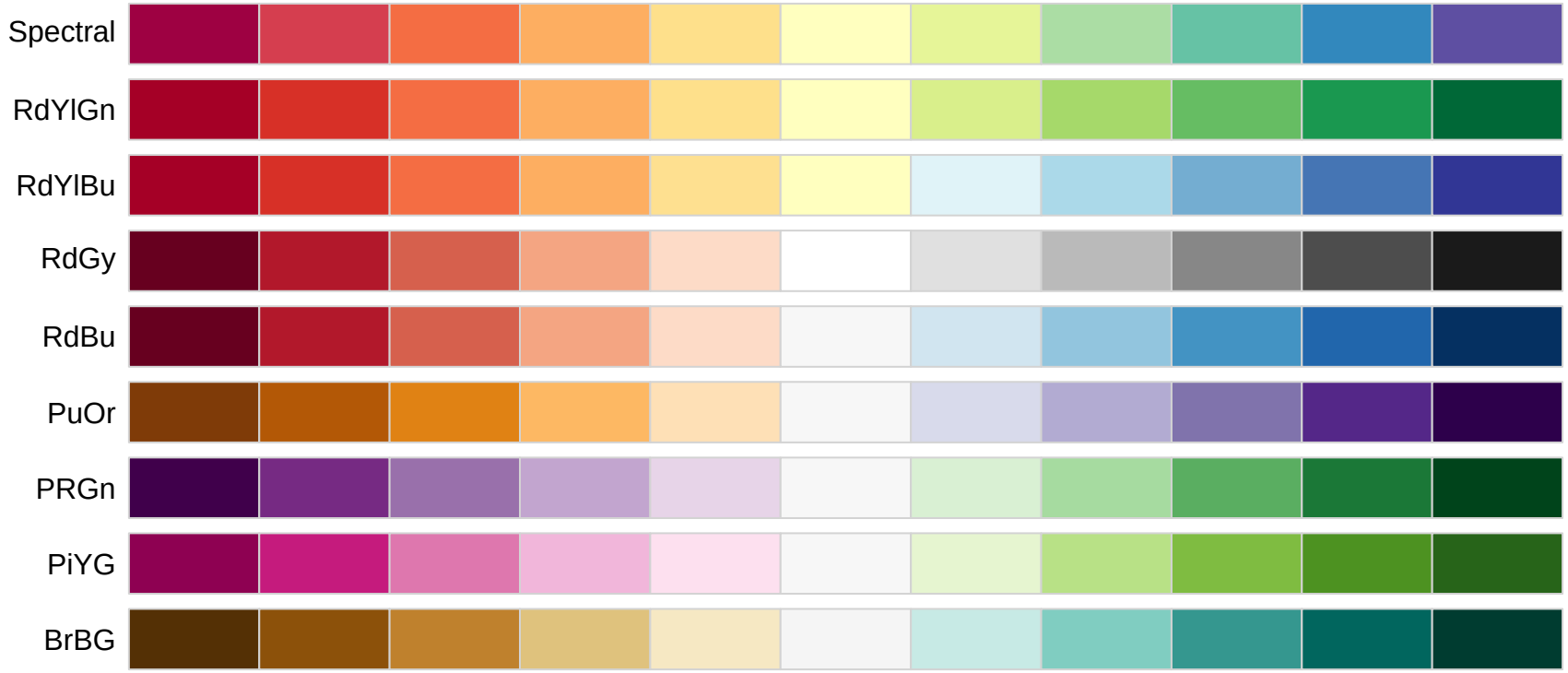


Sequential pallete

Figure 1: ที่มา : สิวะโชติ ศรีสุทธิยากร (2565)

Color to represent data values: Diverging palette

```
display.brewer.all(type = "div")
```



การเรียกใช้สีจาก RcolorBrewer

- ผู้วิเคราะห์สามารถเรียกดูจานสีได้จากฟังก์ชัน `display.brewer.all()` หรือ `display.brewer.pal()`
- การเรียกรหัสสีจากแต่ละจานสีขึ้นมา สามารถทำได้ด้วยฟังก์ชัน `brewer.pal(n, name)`

```
brewer.pal(n = 5, name = "RdBu")
```

```
[1] "#CA0020" "#F4A582" "#F7F7F7" "#92C5DE" "#0571B0"
```

```
brewer.pal(n = 5, name = "Dark2")
```

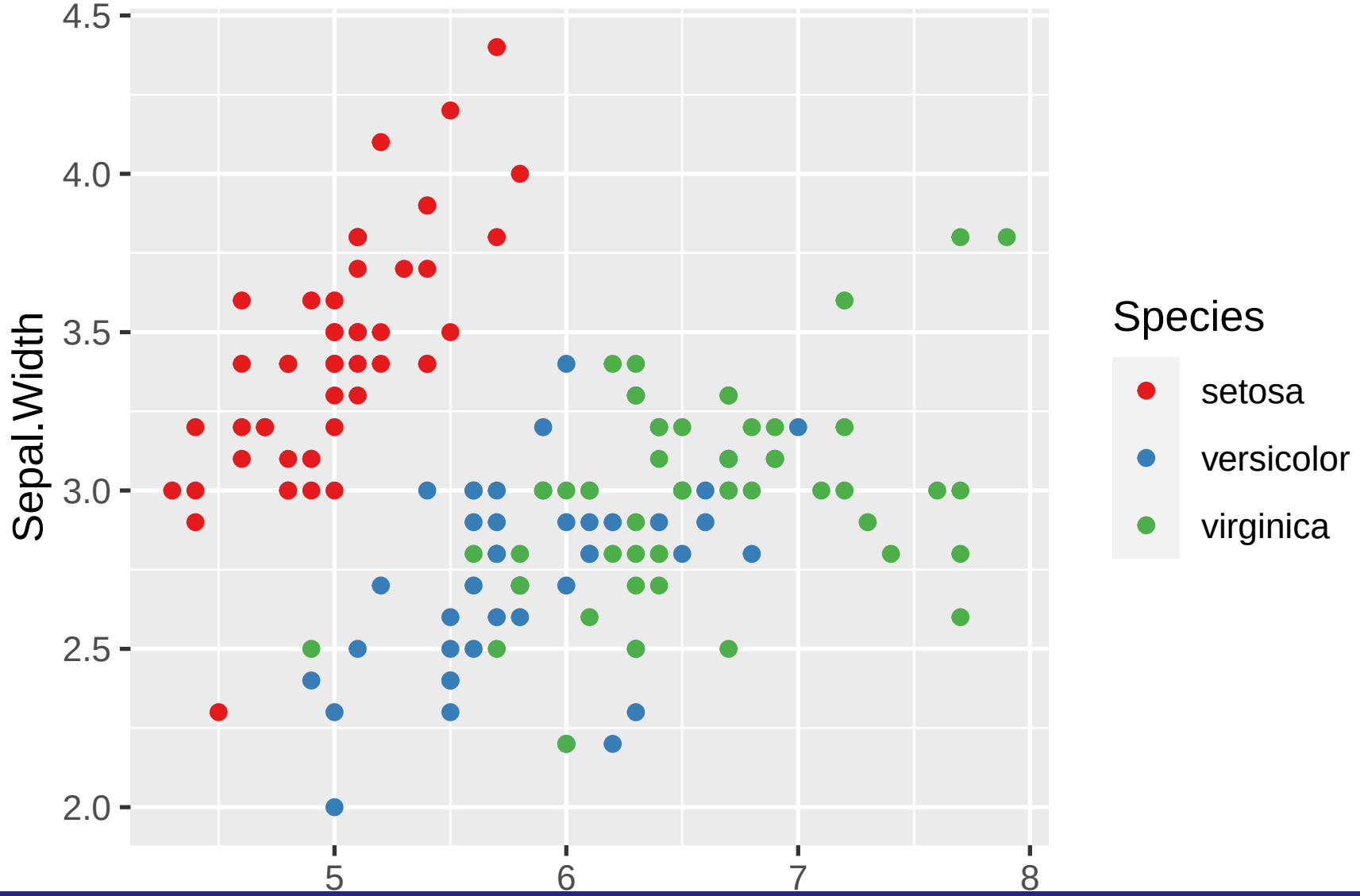
```
[1] "#1B9E77" "#D95F02" "#7570B3" "#E7298A" "#66A61E"
```

การเรียกใช้สีจาก RcolorBrewer

- สำหรับ ggplot2 มีฟังก์ชัน `scale_color_brewer(type, palette, direction)` ที่สามารถเรียกใช้จานสีใน package ดังกล่าวได้ง่ายขึ้น เช่น

```
iris %>% ggplot(aes(x = Sepal.Length,
                    y = Sepal.Width,
                    col = Species))+
  geom_point()+
  scale_color_brewer(type = "qual", palette = "Set1")
```

การเรียกใช้สีจาก RcolorBrewer

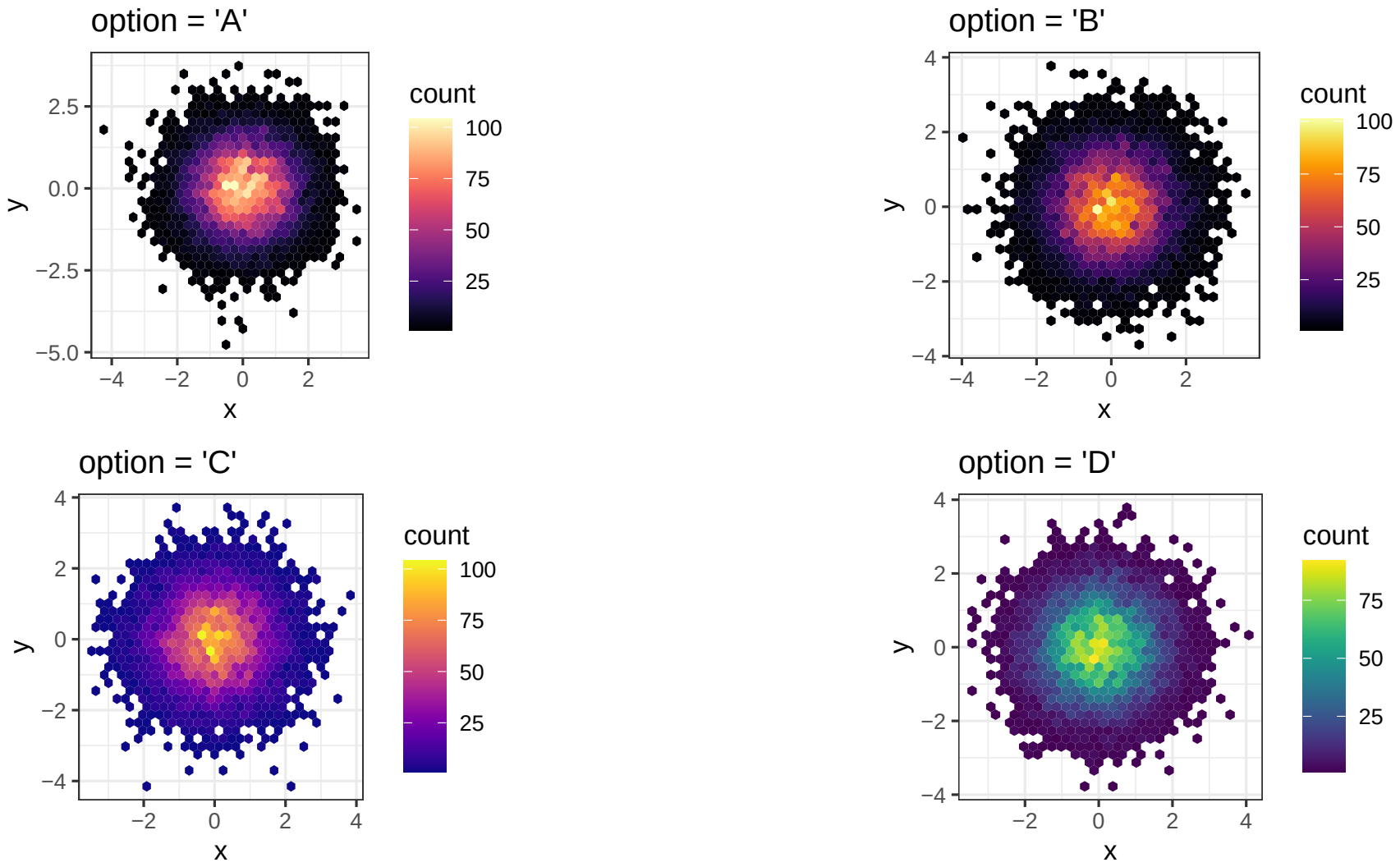


การเรียกใช้สีจาก Viridis color

```
install.packages("viridis")
library(viridis)
```

- scale_color_viridis_d(alpha = 1, begin = 0, end = 1, direction = 1, option = "D")
- scale_color_viridis_c()

การเรียกใช้สีจาก Viridis color



grey scale color

สำหรับสิ่งพิมพ์ที่เป็นขาวดำ จานสีที่เหมาะจะใช้ในทัศนภาพข้อมูลคือ grey scale color โดยสามารถกำหนดผ่านฟังก์ชัน `scale_color_grey()` โดยผลลัพธ์ที่ได้เป็นดังนี้ (สำหรับตัวแปรต่อเนื่องให้ใช้ `scale_color_gradient()` แทน)

```
iris %>%
  ggplot(aes(x = Sepal.Length,
             y=Sepal.Width,
             col=Species))+
  geom_point()+
  scale_color_grey(start = 0.2, end = 0.8)+
  theme_classic()
```

Color as highlight tool

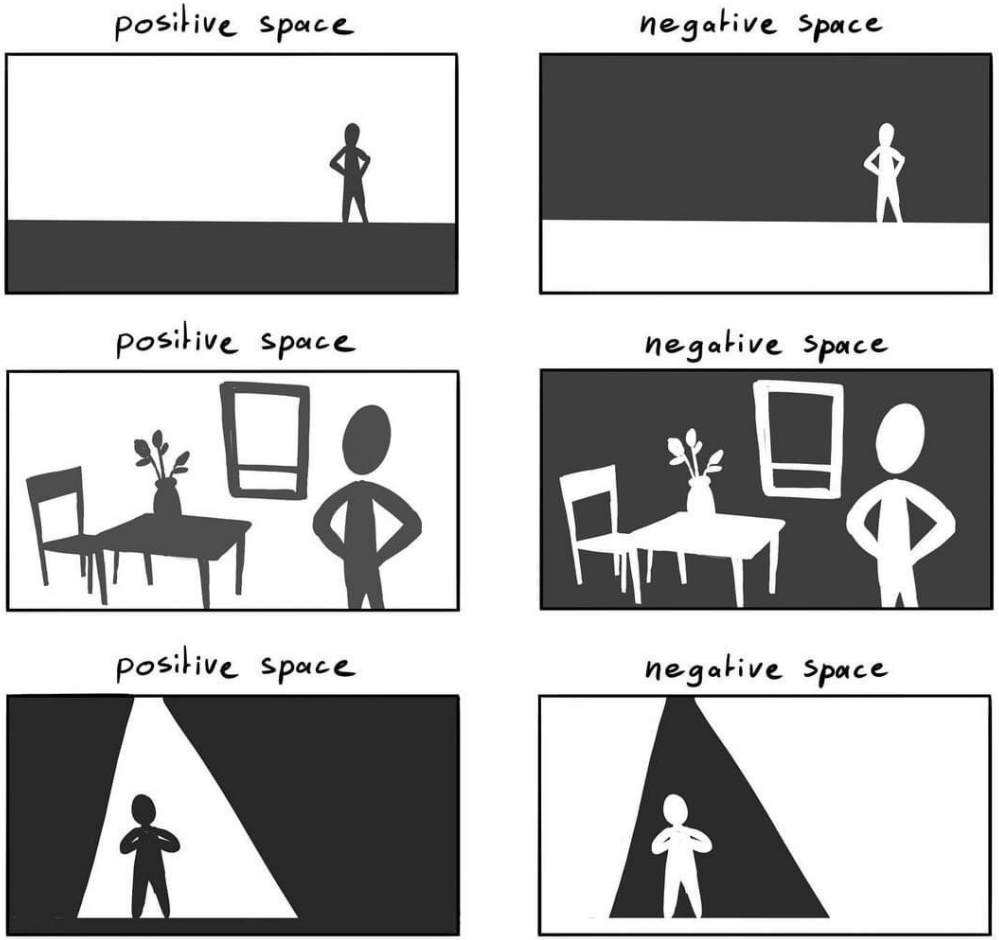
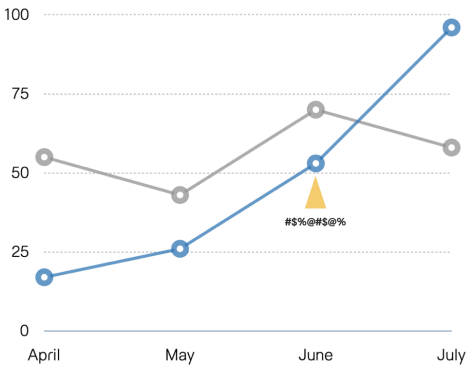


Figure 2: ที่มา : facebook

Color as highlight tool

- ใช้สีความเข้มต่ำเป็นพื้นหลัง
- ใช้สีความเข้มสูงบนใจความสำคัญ
- ไม่ควรใช้การไล่สีหรือเงาบนวัตถุ/ตัวอักษร

Background	Main	Accent
70	25	5



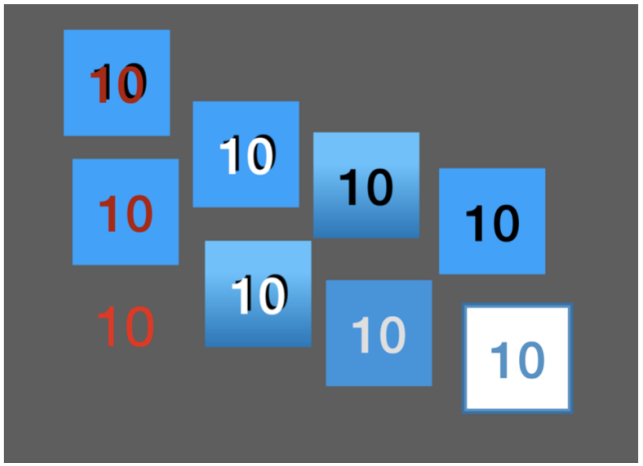
Color as highlight tool

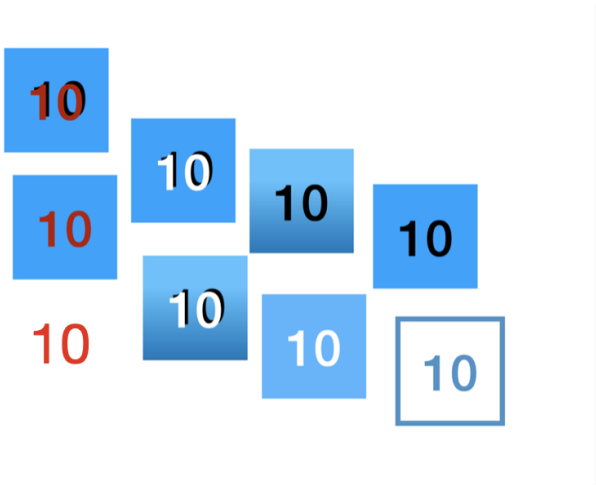
Color is Matter!!

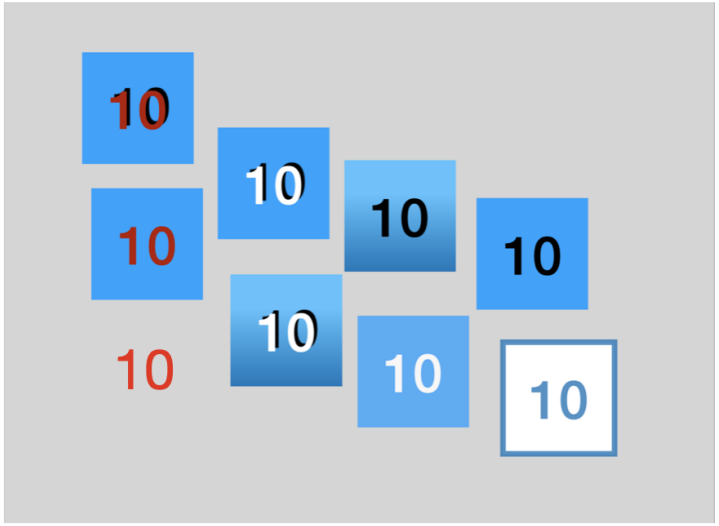
Color as highlight tool

- Color can also be an effective tool to highlight specific elements in the charts.
- There may be specific categories or values in the dataset that carry main idea/information about the story we want to tell.

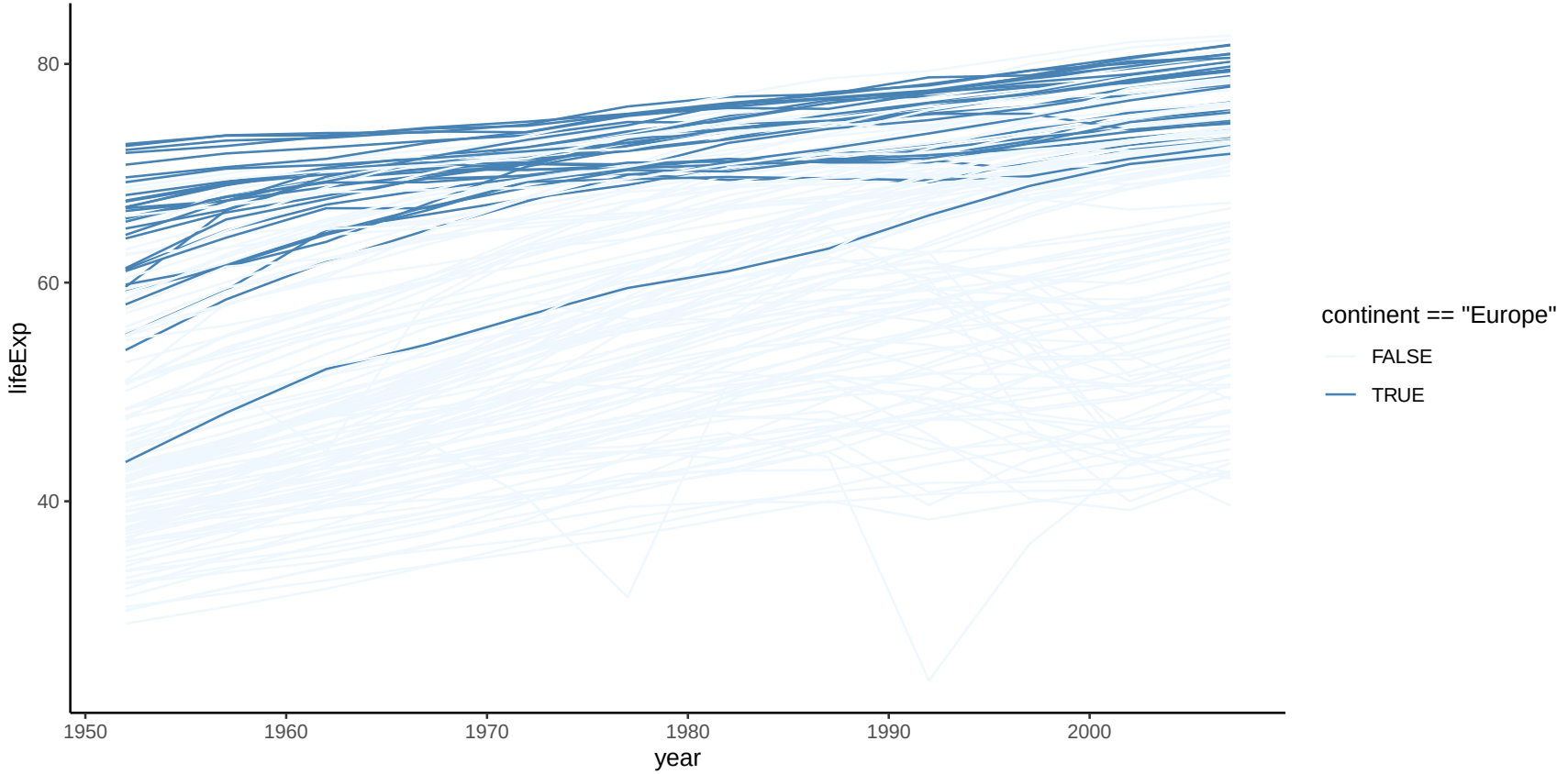
Color as highlight tool



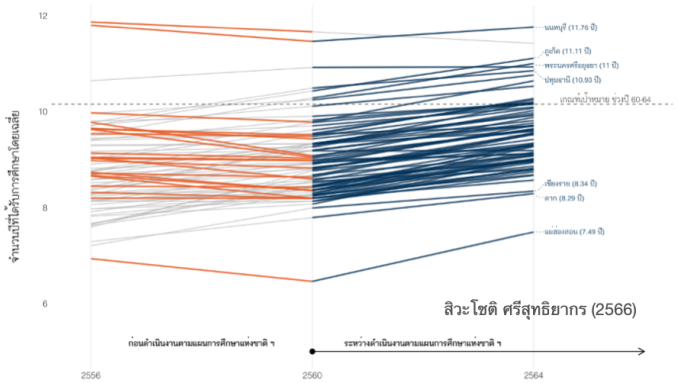


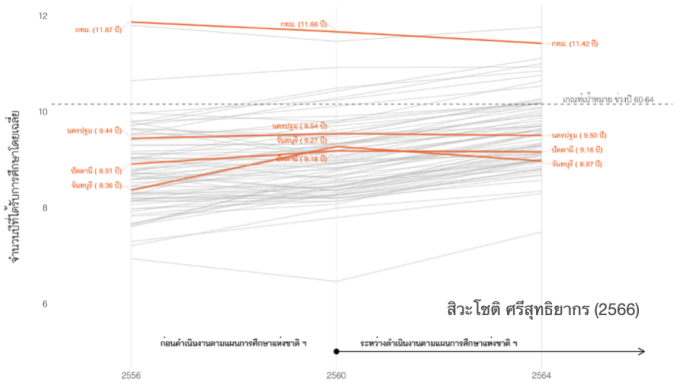


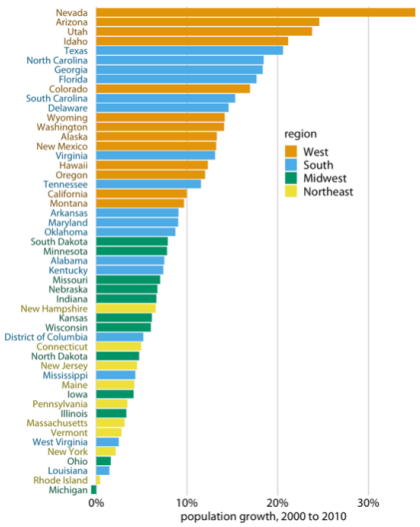
Color as highlight tool

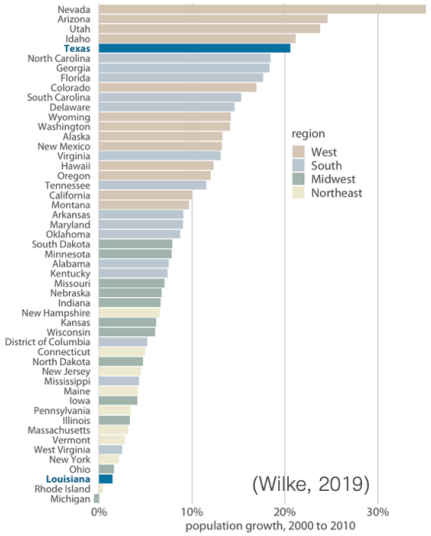


Color as highlight tool

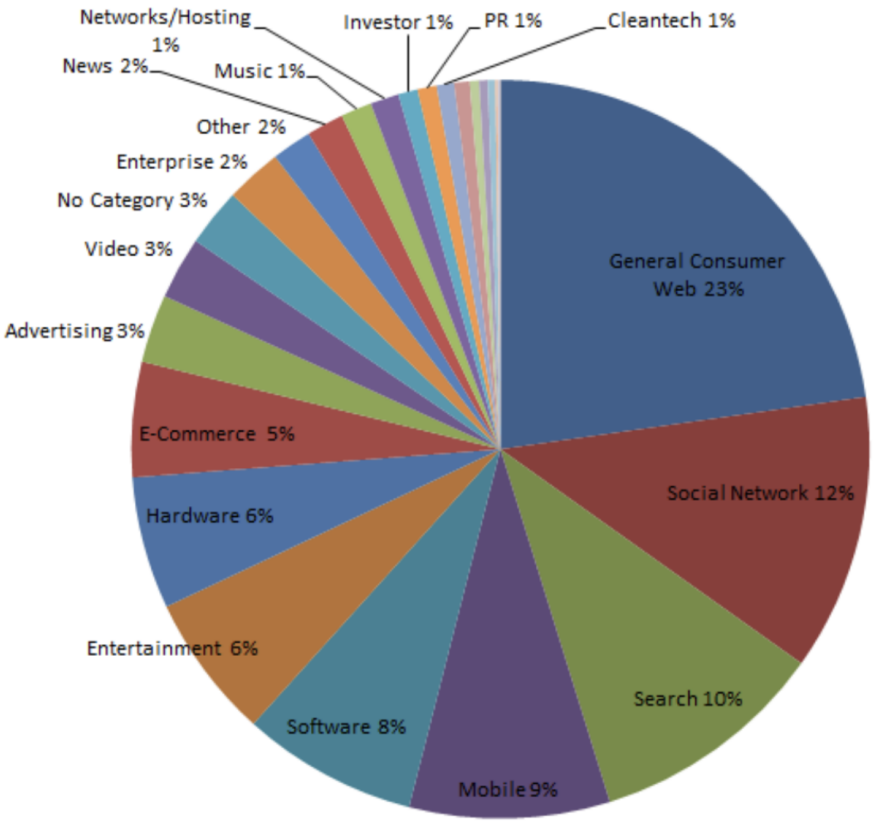




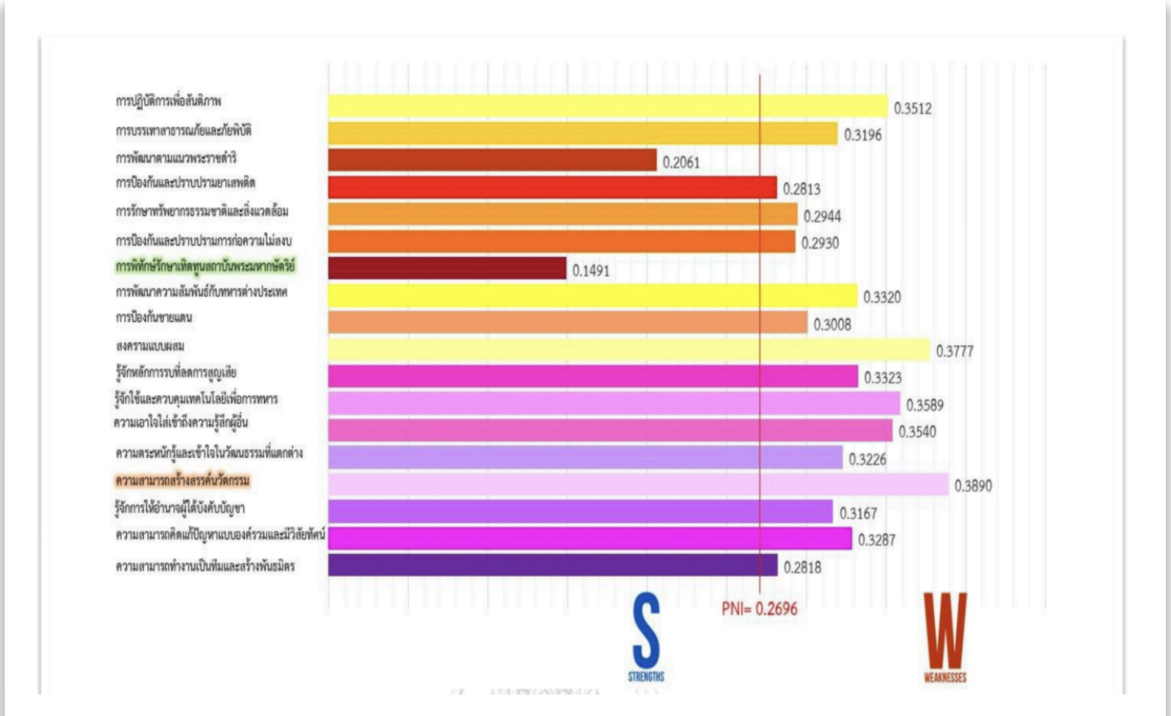




Useless color

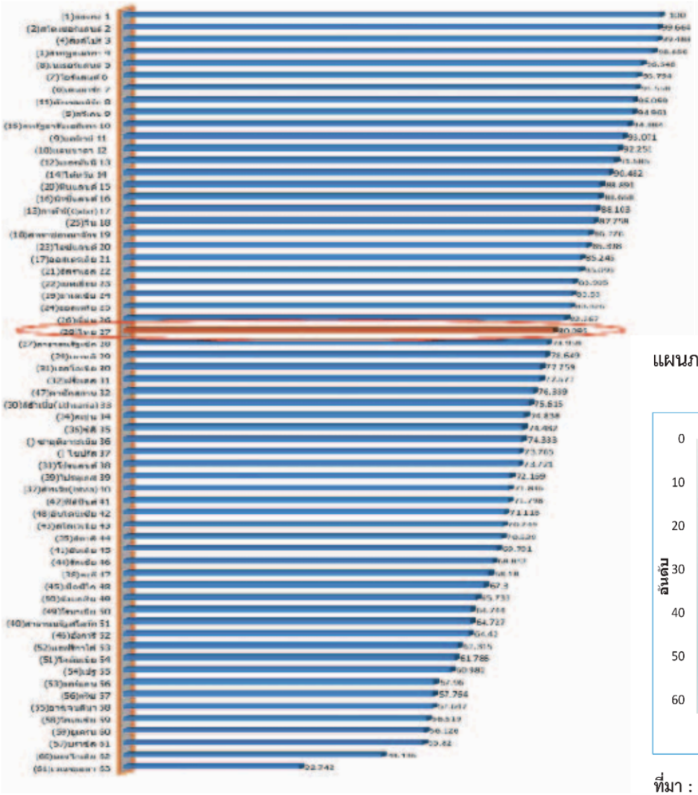


Useless color



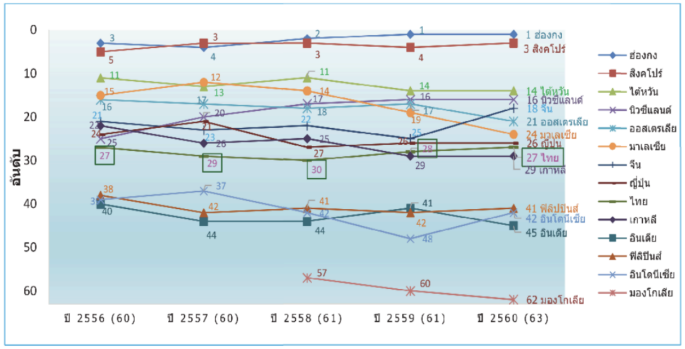
Useless color

แผนภาพ 2 ภาพรวมในการแข่งขันของนานาประเทศ ปี 2560



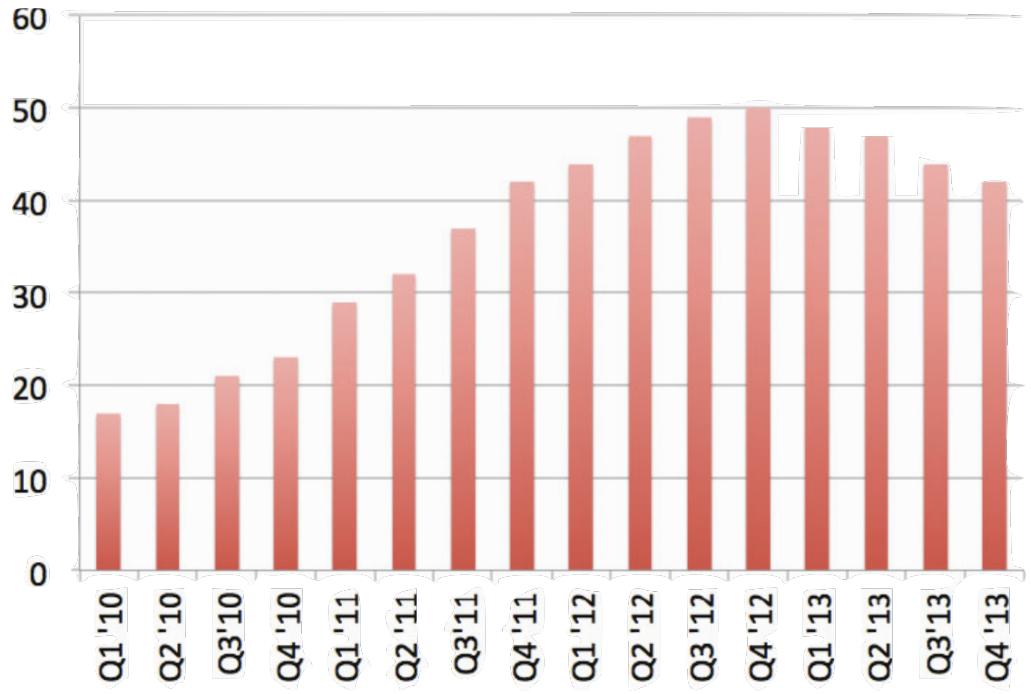
ที่มา : IMD World Competitiveness Yearbook 2017
(ตัวเลขในวงเล็บแสดงอันดับ ปี 2559)

แผนภาพ 3 แนวโน้มสมรรถนะในภาพรวมของประเทศไทย
เปรียบเทียบกับประเทศในภูมิภาคเอเชียแปซิฟิก ปี 2556 – 2560



ที่มา : IMD. World Competitiveness Yearbook 2017

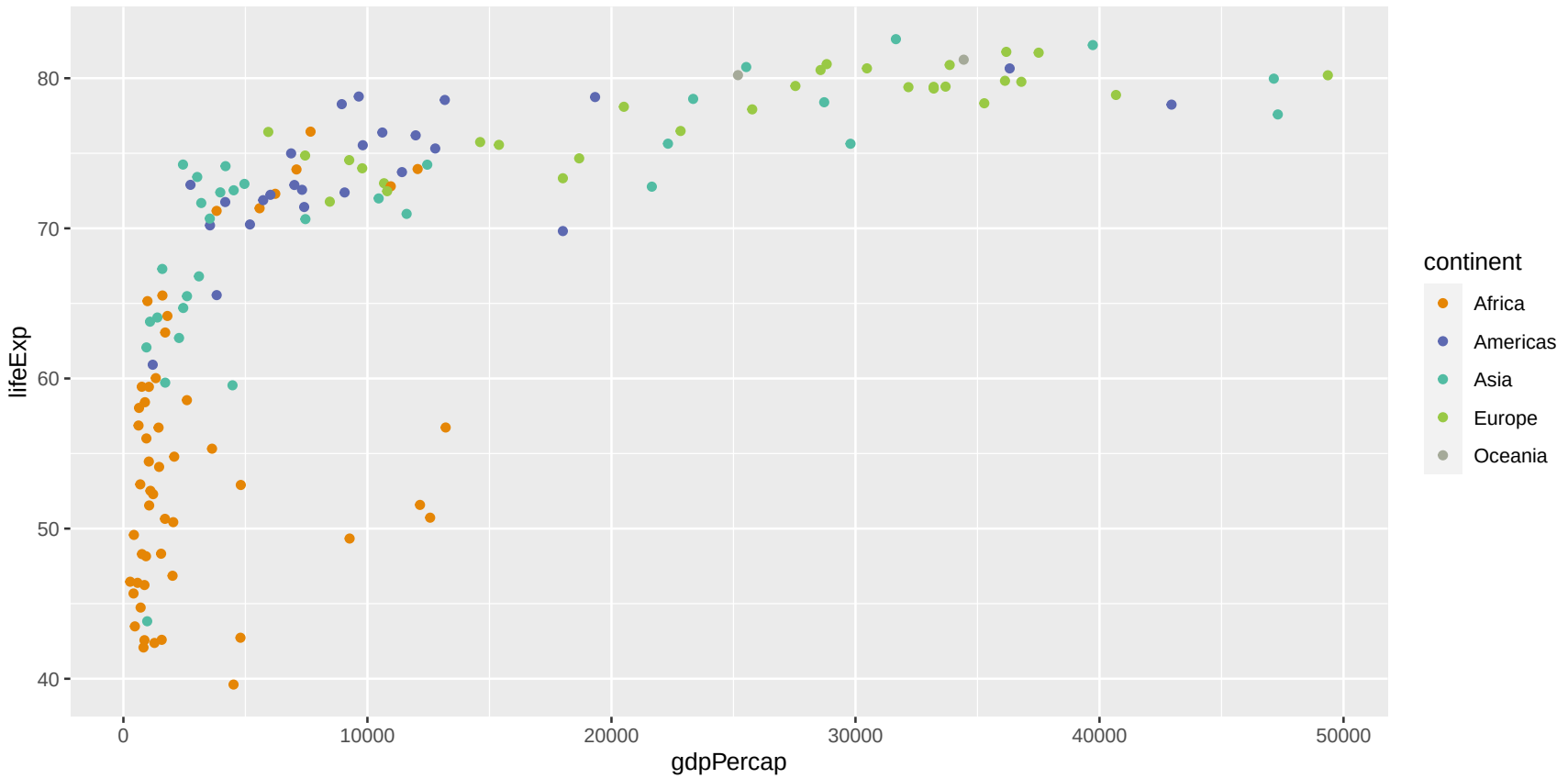
Useless color



กิจกรรม : Color is Matter

จากแผนภาพการกระจายระหว่าง lifeExp กับ gdpPercap
ขอให้ผู้เรียนทดลองกำหนดสีให้กับจุดบนแผนภาพ โดยทดลองใช้สีให้ครบทั้ง 3 วัตถุประสงค์ และเปรียบเทียบสารสนเทศที่ได้ว่ามีความแตกต่างกันอย่างไร

สรุปผลลัพธ์ที่ได้ในขั้นตอนนี้



Aesthetic mapping “size” and “shape”

กิจกรรม aesthetic mapping: size

จากแผนภาพการกระจายอันล่าสุด ขอให้ผู้เรียนทดลองกำหนด aesthetic mapping ของ size โดยจับคู่กับ จำนวนประชากรของประเทศ **ผลลัพธ์ที่ได้เป็นอย่างไร?**

Syntax เดิม

```
gapminder %>% filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y = lifeExp,
           col = continent))+
      geom_point()+
  scale_color_viridis_d(option="A", begin=0, end = 0.8)
```

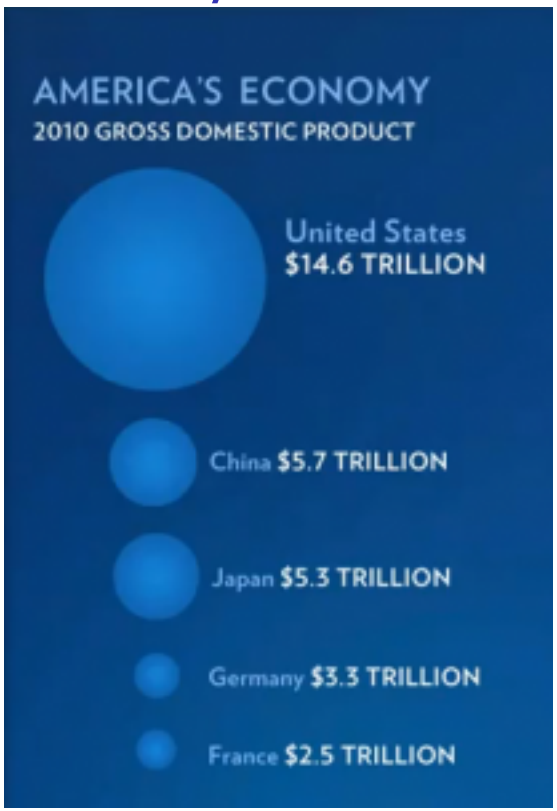
Size

In the 2010 Gross Domestic Product chart, the radius of each circle has been scaled based on the size of each nation’s economy.

Figure 3: ที่มา : https://www.youtube.com/watch?v=kI2g40GoRxg

Size

Scaled by radius



Scaled by area

ที่มา : https://blog.fastfedora.com/2011/01/2011-state-of-the-union-visualizations.html

scale_size_XXX()

- `scale_size(name, breaks, labels, range=c(1,6))`
- `scale_radius(name, breaks, labels, range=c(1,6))`
- `scale_size_binned(..., n.breaks)`
- `scale_size_area(..., max_size = 6)`
- `scale_size_binned_area(..., max_size = 6)`

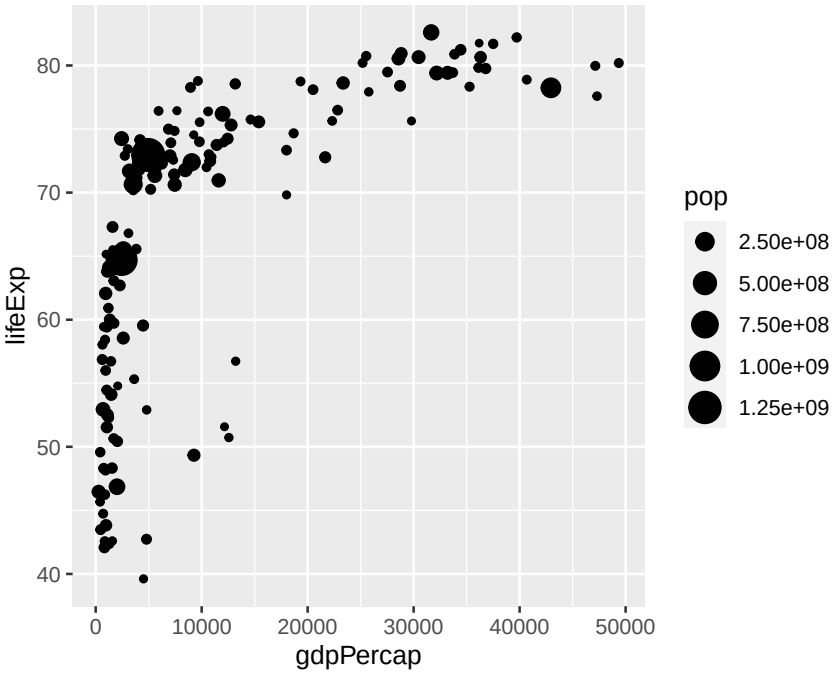
scale_size_XXX()

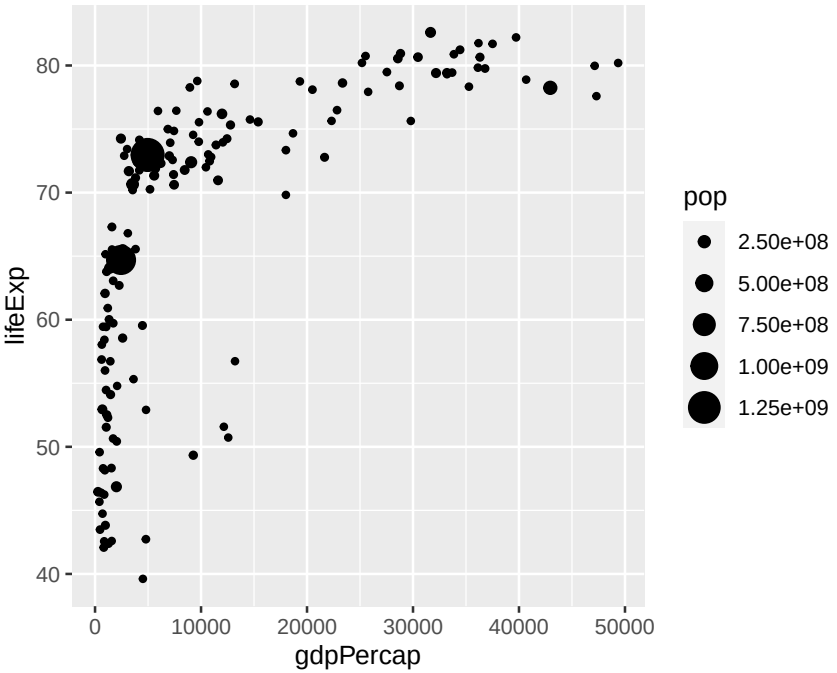
```
gapminder %>% filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y = lifeExp,
             size = pop))+
  geom_point()+
  scale_size()

gapminder %>% filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y = lifeExp,
             size = pop))+
  geom_point()+
  scale_radius()
```

scale_size_XXX()

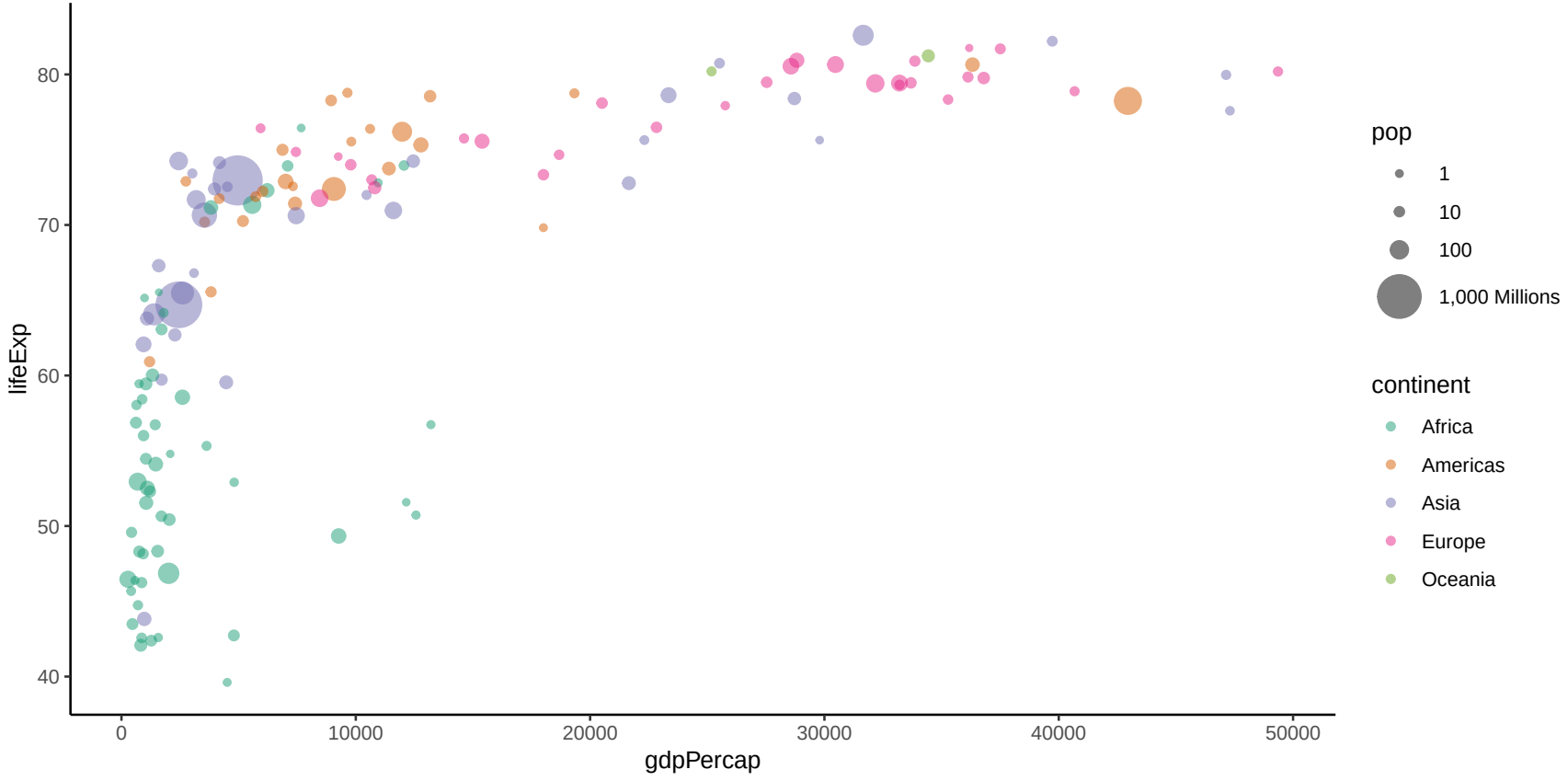
ผลลัพธ์ทั้งสองมีความแตกต่างกันอย่างไร ?



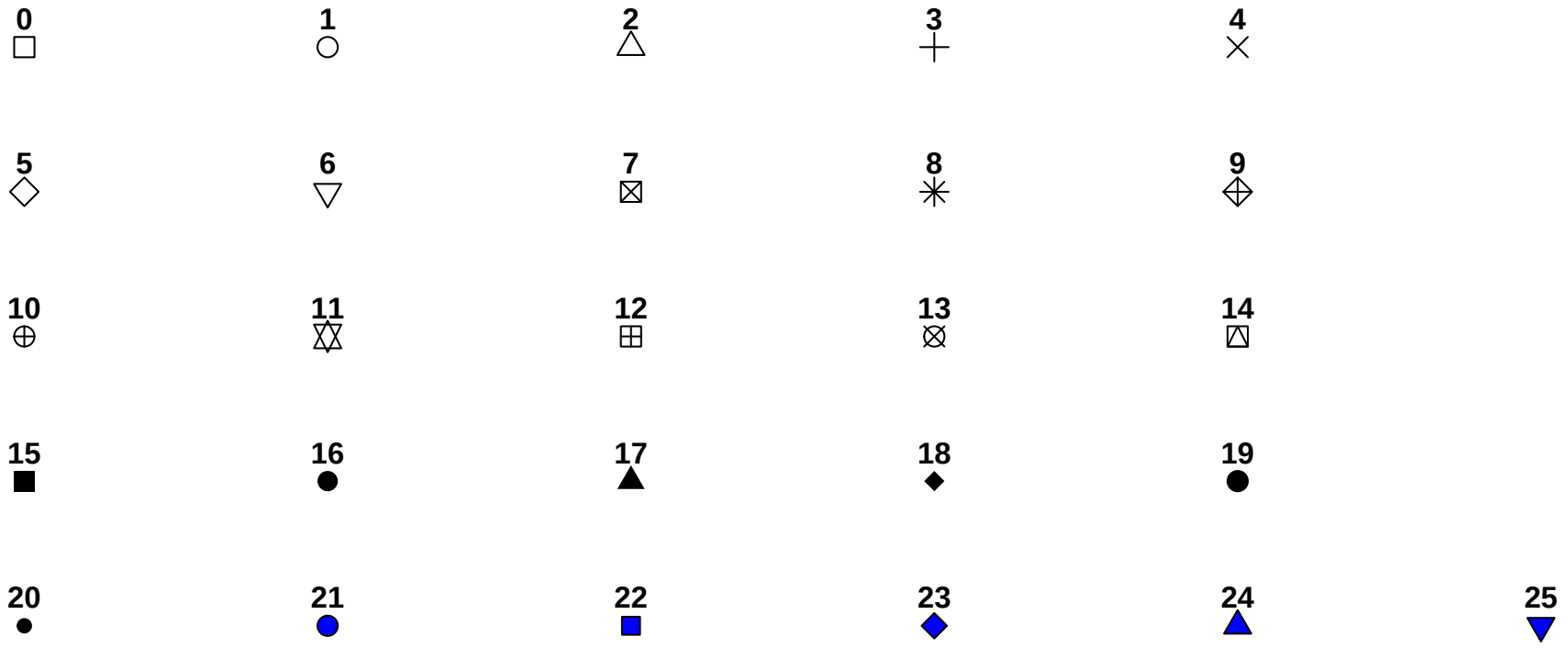


กิจกรรม : scale_size()

ขอให้ผู้เรียนปรับแต่ง size ของจุด ให้เป็นเหมือนกับในรูปตัวอย่างด้านล่าง



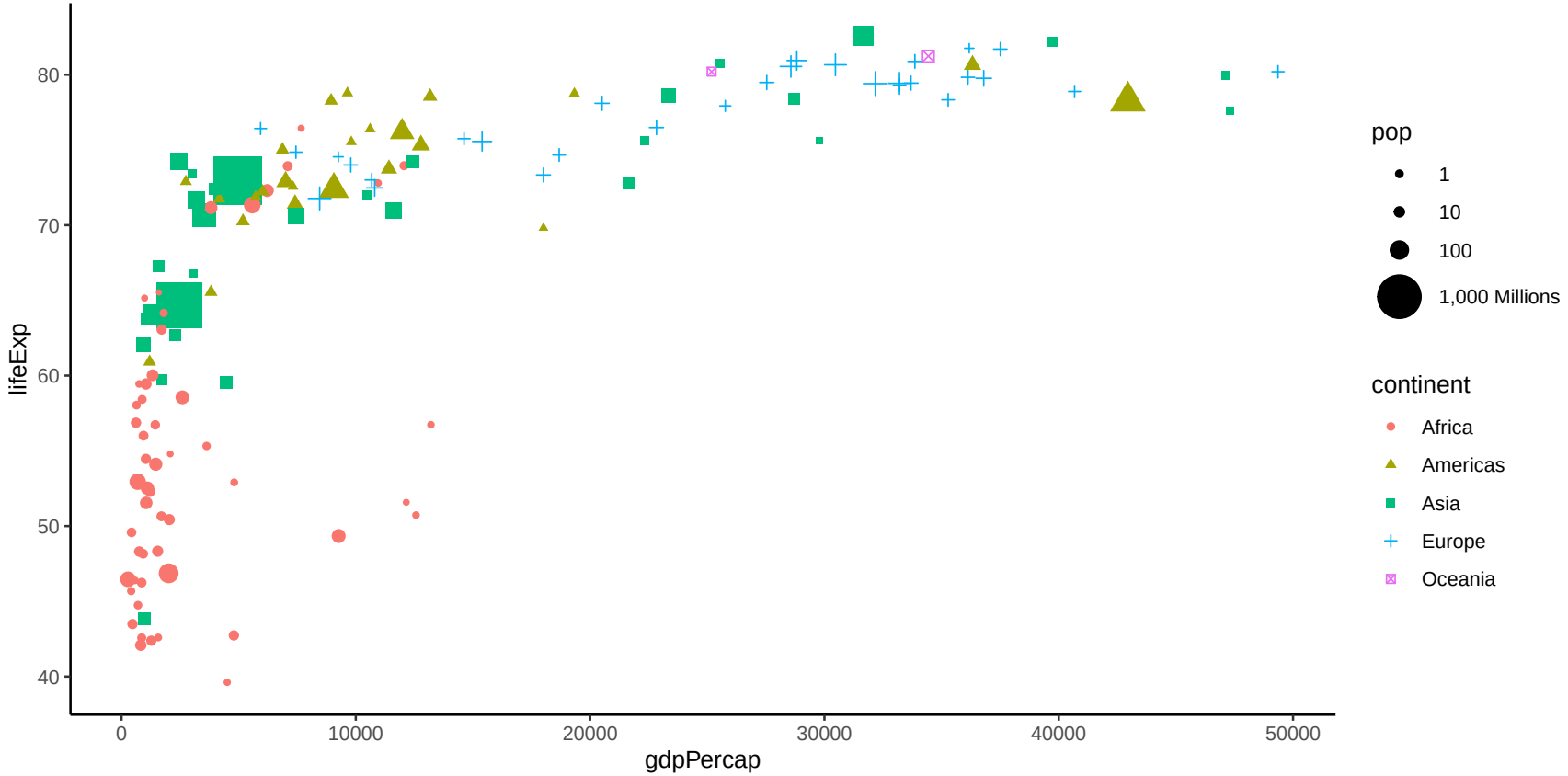
Shape of point



Shape

```
gapminder %>% filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y = lifeExp,
             size = pop,
             col = continent,
             shape = continent))+
  geom_point()+
  theme_classic()
```

Shape

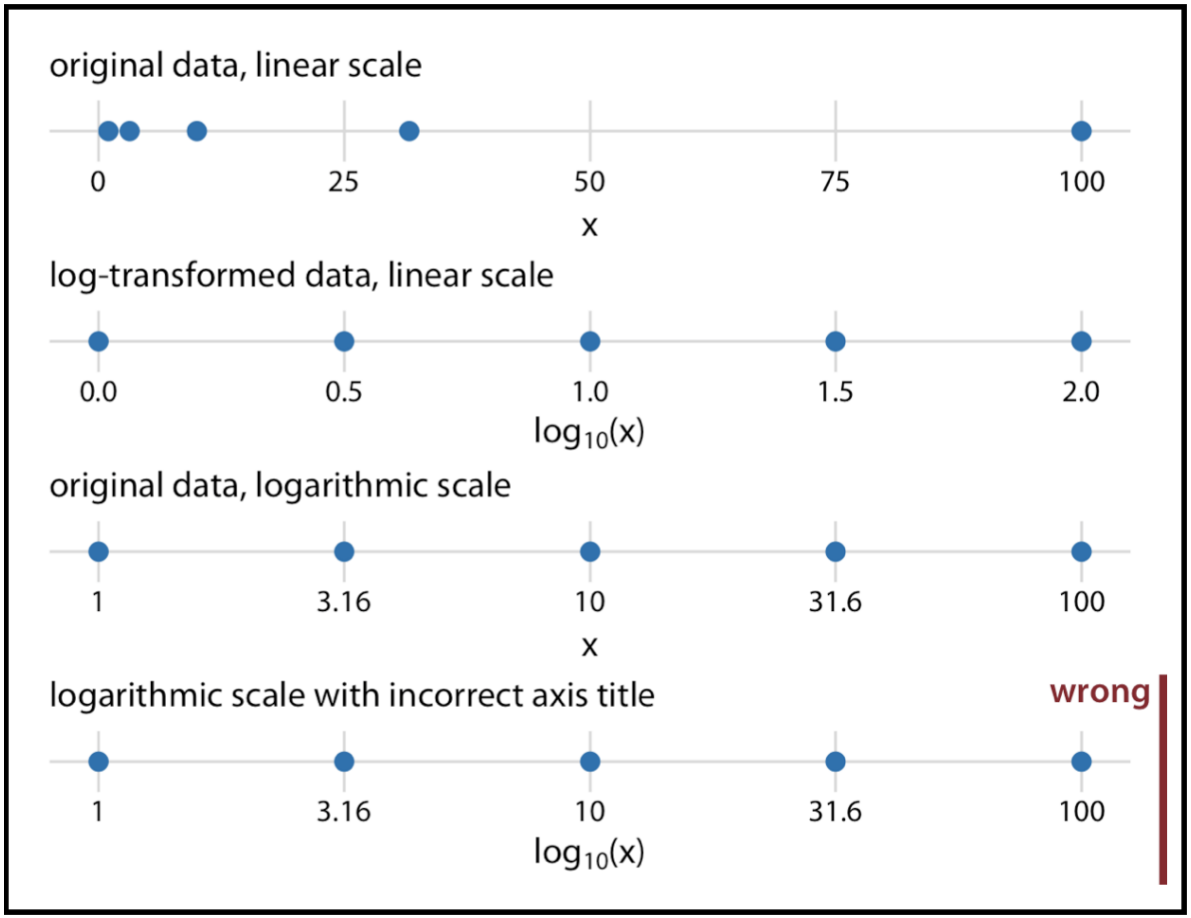

lifeExp
80
70
60
50
40
0
10000
20000
30000
40000
50000
gdpPercap
pop
1
10
100
1,000 Millions
continent
Africa
Americas
Asia
Europe
Oceania

scale_shape_XXX()

- `scale_shape(breaks, labels)`
- `scale_shape_binned(breaks, labels)`
- `scale_shape_manual(values)`

scale of the axes

Linear vs Nonlinear axes



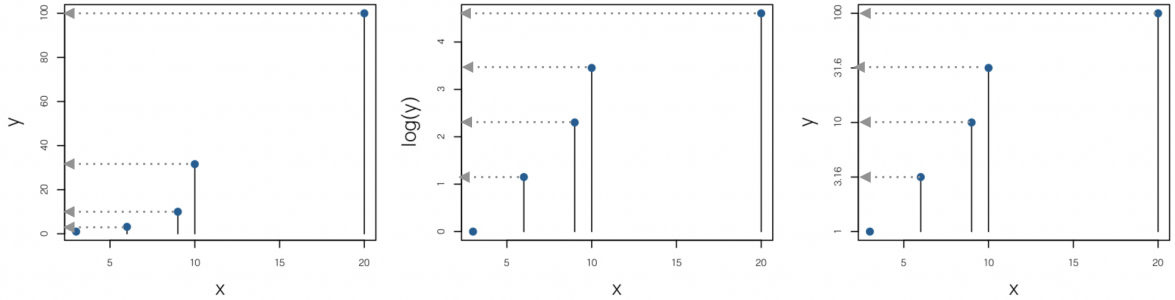
ที่มา : Wike (2019)

Linear Scale

- `scale_x_continuous(name, breaks, labels, limits)`
- `scale_y_continuous()`
- `scale_x_discrete(name, limits, labels)`
- `scale_y_discrete()`

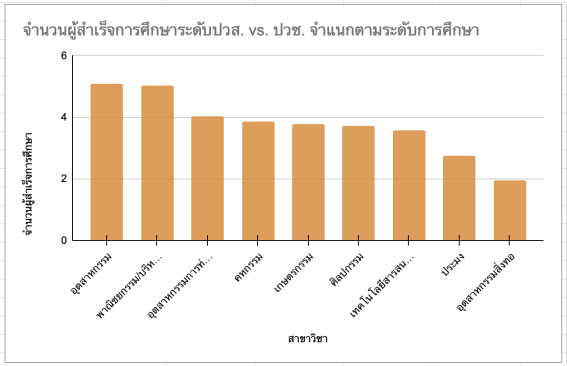
Logarithmic Scale

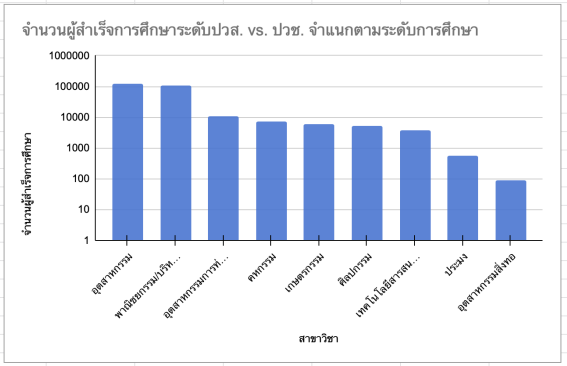
- `scale_x_log10(name, limits, labels)`
- `scale_y_log10()`



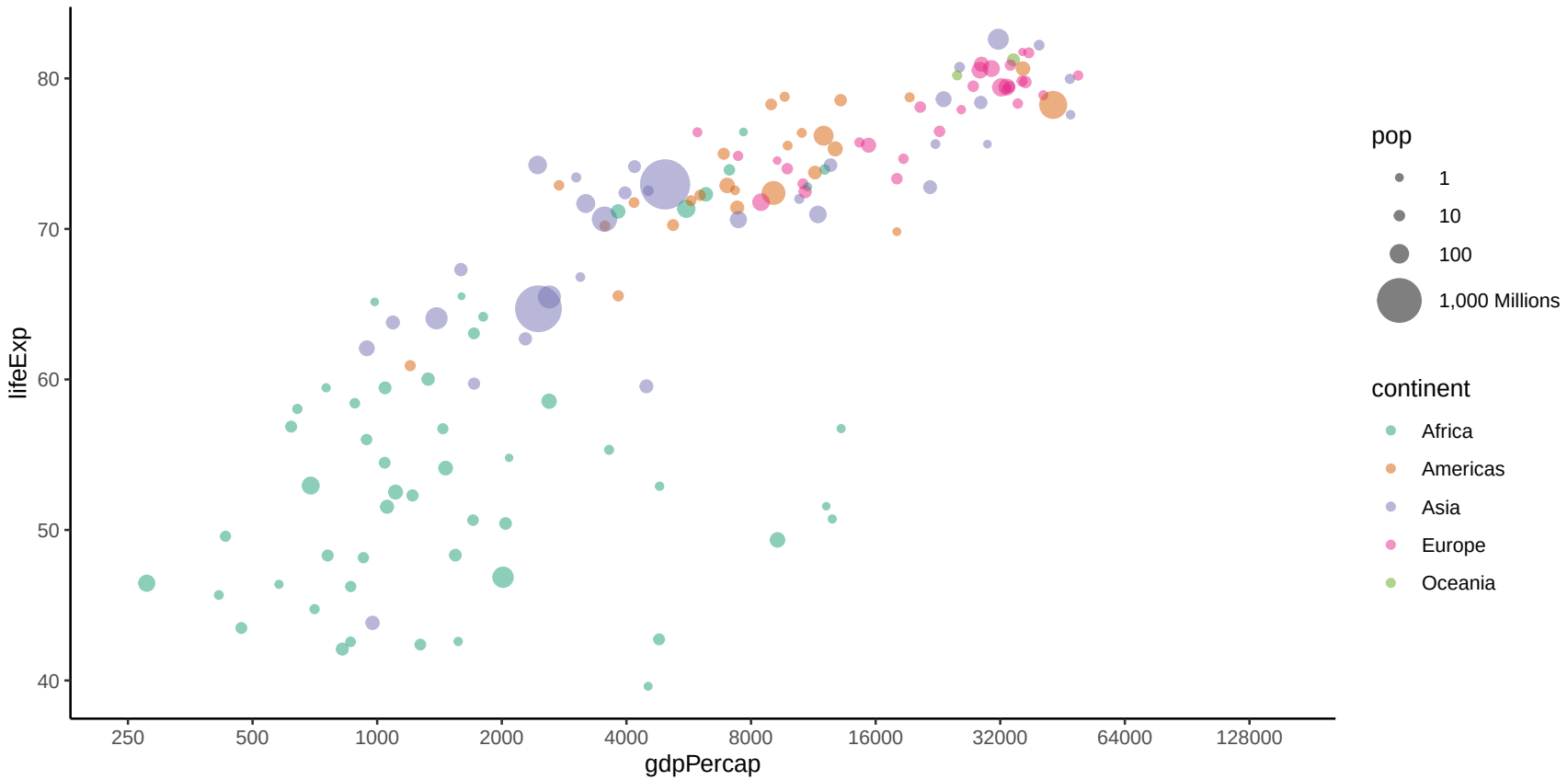
สาขาวิชา	รวม	ปวช.	ปวส.	log(รวม)
อุตสาหกรรม	123268	65724	57544	5.09085035
พาณิชยกรรม/บริหารธุรกิจ	106556	55257	51299	5.027577909
อุตสาหกรรมการท่องเที่ยว	10779	6696	4083	4.032578472
คหกรรม	7171	4101	3070	3.855579723
เกษตรกรรม	6026	2926	3100	3.780029127
ศิลปกรรม	5175	4163	1012	3.713910354
เทคโนโลยีสารสนเทศฯ	3702	1758	1944	3.568436414
ประมง	576	91	485	2.760422483
อุตสาหกรรมสิ่งทอ	87	39	48	1.939519253

จำนวนผู้สำเร็จการศึกษาระดับ ปวช. ปวส. และอนุปริญญา ในปีการศึกษา 2562 จำแนกตามระดับการศึกษา





กิจกรรม


lifeExp
80
70
60
50
40
250
500
1000
2000
4000
8000
16000
32000
64000
128000
gdpPercap
pop
1
10
100
1,000 Millions
continent
Africa
Americas
Asia
Europe
Oceania

Adding data label

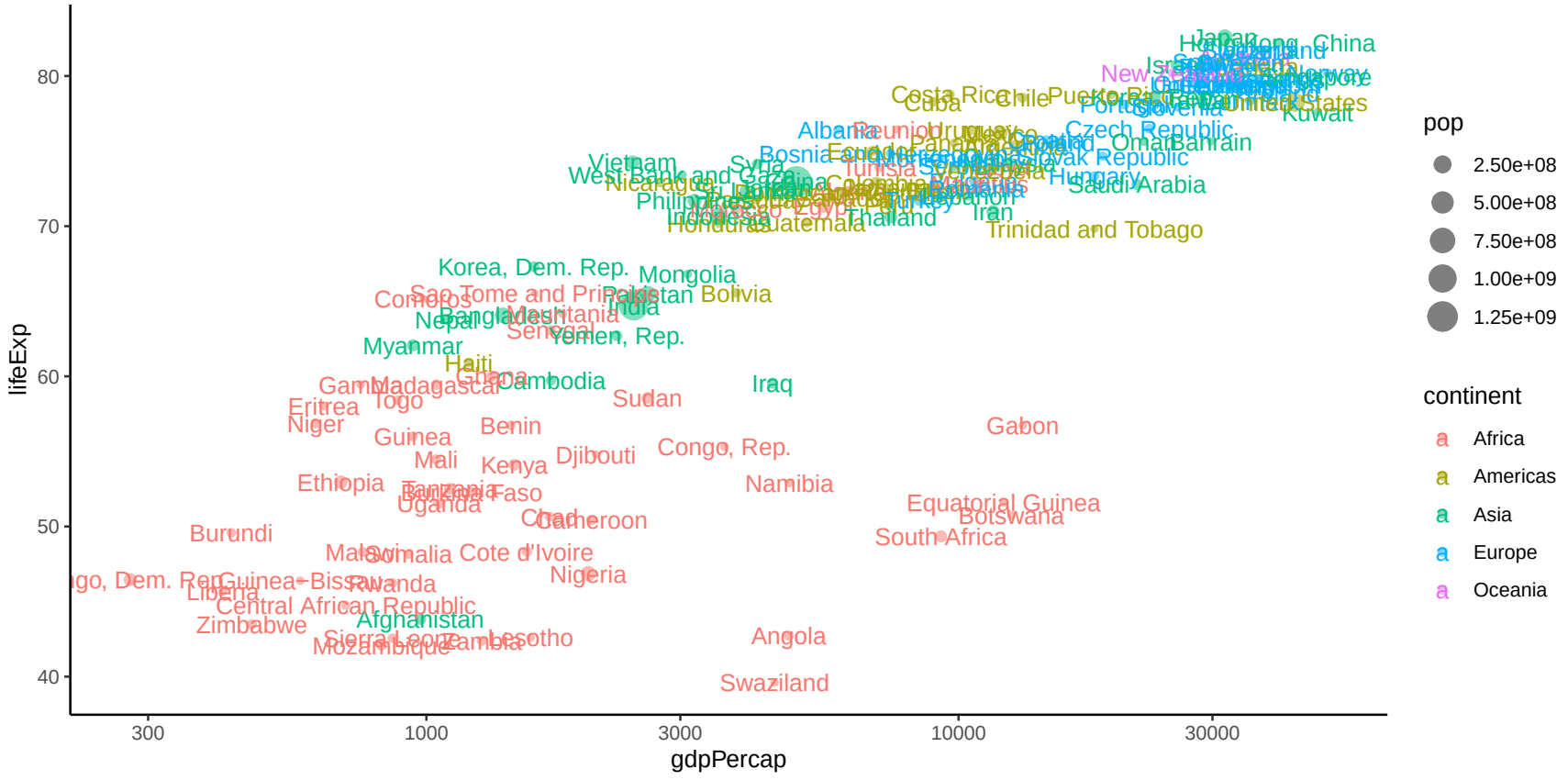
Adding data label using geom_text()

ฟังก์ชัน geom_text() เป็น geometric layer ที่ใช้เมื่อต้องการใส่ label ให้กับข้อมูลภายในทัศนภาพข้อมูล โดยมักใช้ร่วมกับ geometric layer ตัวอื่น เช่น geom_point(), geom_bar() เป็นต้น

ตัวอย่างต่อไปนี้แสดงการใช้ geom_text() เพื่อใส่ label ให้กับข้อมูล

ผลลัพธ์ที่ได้เป็นอย่างไร?

```
gapminder %>%
  filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y= lifeExp,
             col=continent))+
  geom_point(aes(size = pop),alpha=0.5)+
  geom_text(aes(label = country))+
  scale_x_log10()
```


lifeExp
80
70
60
50
40
300
1000
3000
10000
30000
gdpPercap
pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09
continent
Africa
Americas
Asia
Europe
Oceania

geom_text()

ฟังก์ชัน geom_text() มีอาร์กิวเมนท์สำคัญดังนี้

```
geom_text(data = NULL,
          mapping = aes(...),
          nudge_x = 0, nudge_y = 0,
          check_overlap = FALSE,
          show.legend = NA,
          family,
          angle)
```

การกำหนดอาร์กิวเมนท์ดังกล่าว ทำให้ได้ทัศนภาพข้อมูลที่เหมาะสมหรือไม่

Adding data label using geom_text_repel()

```
install.packages("ggrepel")
library(ggrepel)
gapminder %>%
  filter(year == "2007") %>%
  ggplot(aes(x = gdpPercap,
             y= lifeExp,
             col=continent))+
  geom_point(aes(size = pop),alpha=0.5)+
  geom_text_repel(aes(label = country))+
  scale_x_log10()+
  theme_classic()
```

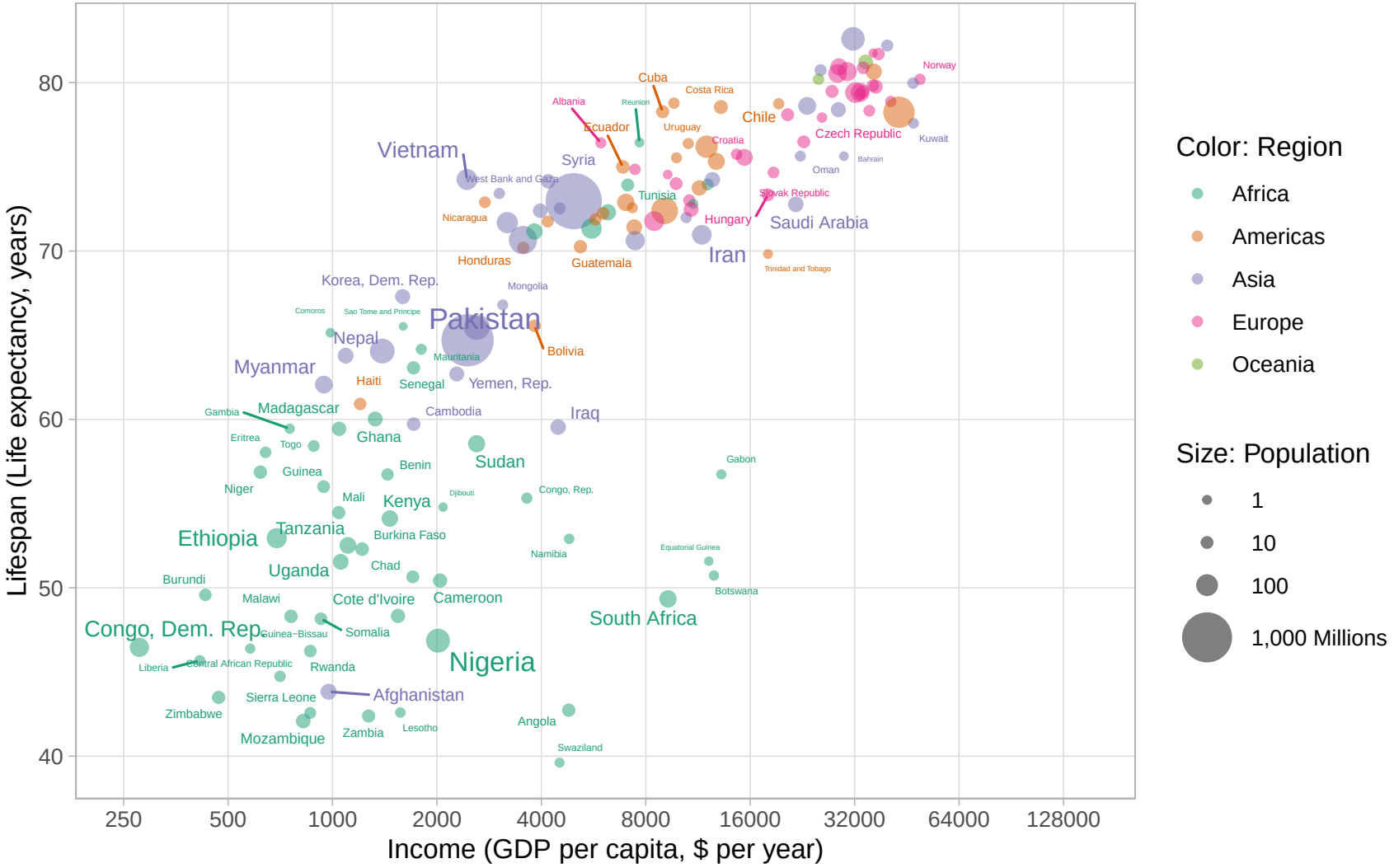
geom_text_repel()

```
geom_text_repel(data = NULL,
                mapping = aes(...),
                nudge_x = 0, nudge_y = 0,
                label.r,
                label.size,
                min.segment.length = 0.5,
                max.overlaps = 10,
                show.legend = NA,
                seed = NA
                )
```

ผลลัพธ์ที่ได้จาก `geom_text()` กับ `geom_text_repel()` มีความเหมือนหรือแตกต่างกันอย่างไร

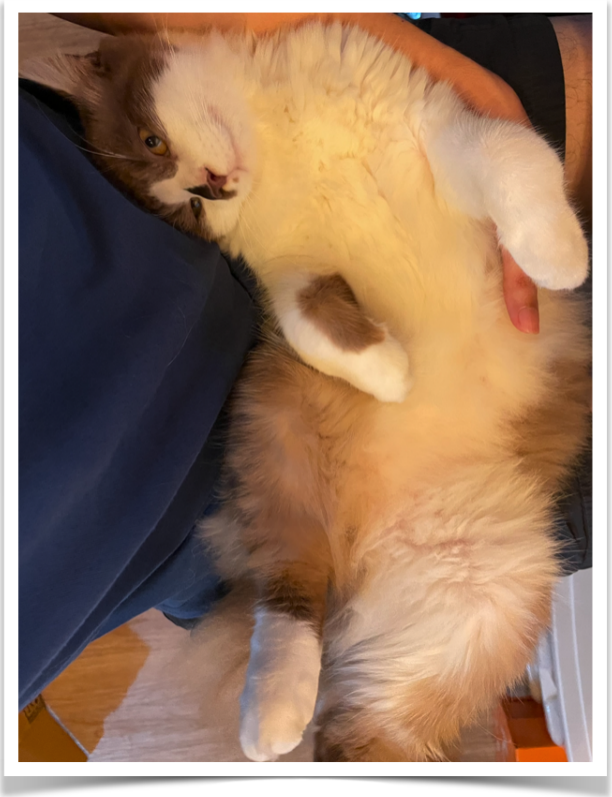
กิจกรรมสรุป

Put it all together



Keep in mind

Qualitative Nominal	Qualitative Ordinal	Quantitative Interval, Ratio
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Containment	Pattern (Texture)	Size (Volume)
Pattern (Density)	Connection	Pattern (Density)
Colour (Lightness)	Containment	Colour (Lightness)
Symbol	Size (Length)	Colour (Hue)
Size (Length)	Angle	Pattern (Texture)
Angle	Size (Area)	Connection
Size (Area)	Size (Volume)	Containment
Size (Volume)	Symbol	Symbol

ที่มา: Mackinlay (1986)

Q & A