

# Data Analysis and Visualization

## Using Stata

เฉลิมพงษ์ คงเจริญ

ภาณุศาสตร์ มนุษยศาสตร์

9 ก.พ. 2568

## Contents

1. การใช้ STATA เป็นอย่างดี	1
การเริ่มใช้ Stata	1
Keeping Track of your STATA session	3
Getting Help in Stata	4
Exploratory Data Analysis in Stata	5
Create new variables in Stata	7
Graphing in Stata	9
2. กราฟสำหรับตัวแปรต่อเนื่องตัวเดียว (ONE CONTINUOUS VARIABLES)	10
Histogram	11
Density Plot	19
Multiple Density Plot	26
Ridgeline Plot (Joy Plots)	29
Cumulative Density Plot	31
Spike Plot	33
Box Plot	36

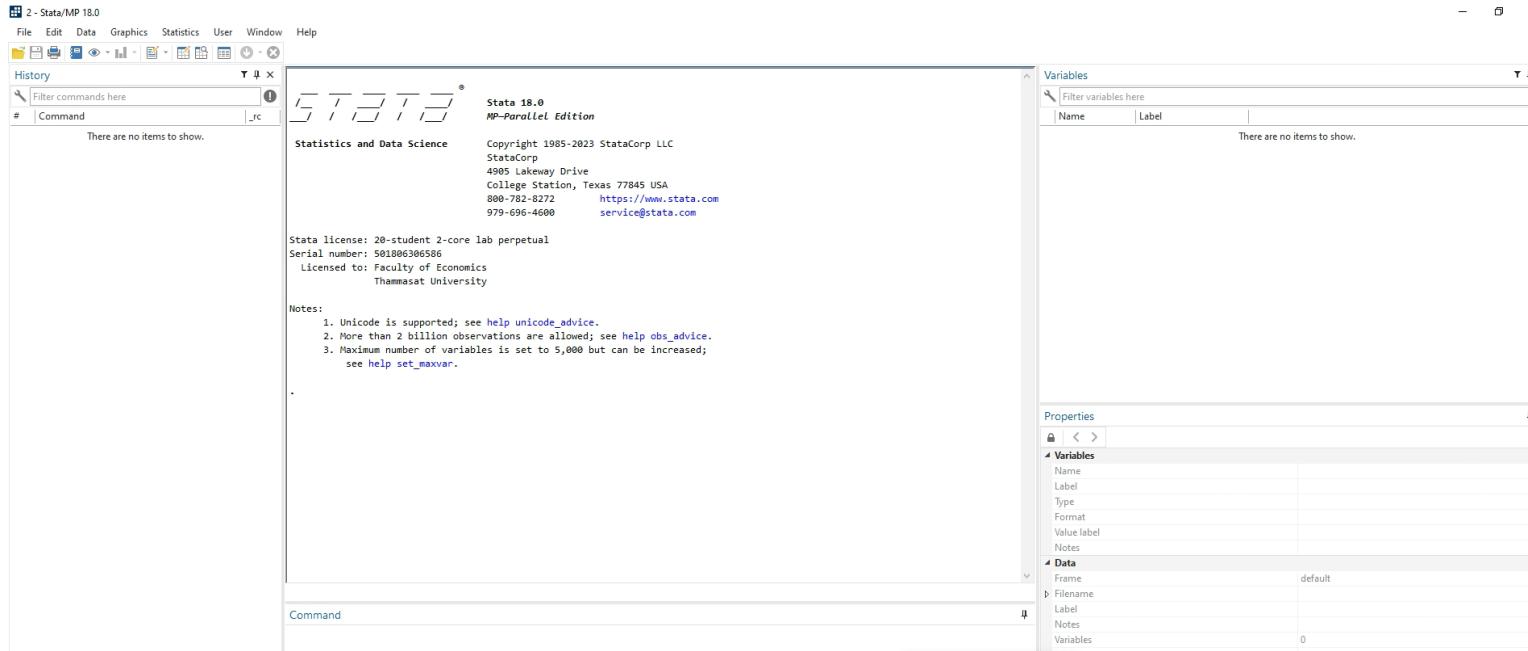
Symmetry Plot .....	43
Quantile-Normal Plot.....	45
Quantile-Quantile (QQ) Plot.....	46
แผนภูมิเท่ง.....	51
แผนภูมิวงกลม .....	58
Dot Chart .....	63
Radar Plot (Spider Plot) .....	68
<b>5. กราฟสำหรับตัวแปรต่อเนื่องสองตัว (Two CONTINUOUS VARIABLES) .....</b>	<b>72</b>
Scatter Plot.....	73
Heat Map .....	81
Hexbin Plot.....	84
เส้นแนวโน้ม .....	86
กราฟเส้น.....	91
แผนภาพพื้นที่ .....	94
<b>6. กราฟสำหรับตัวแปรไม่ต่อเนื่องสองตัว (TWO DISCRETE VARIABLES) .....</b>	<b>96</b>
Jitter Plot.....	97

แผนภูมิตาราง (Table Plot) .....	100
แผนภูมิบล็อก (Block Plot) .....	103
แผนภูมิแท่งซ้อน (Stacked Bar Chart) .....	104
<b>7. กราฟสำหรับตัวแปรสามตัวหรือมากกว่า (THREE OR MORE VARIABLES) .....</b>	<b>108</b>
แผนภูมิ Contour .....	109
แผนภูมิฟองอากาศ (Bubble Plot) .....	112

# 1. การใช้ STATA เป็นองค์น

## การเริ่มใช้ Stata

Stata interface (Dropdown menus, Command windows – it will save time but with some learning curve)



4 windows (Result window, command window, history window, variable window)

- Dropdown menus

Files, Graphics, Statistics, Help

- Command window

- Open data set

Open data .dta

- Import other types e.g. Excel spreadsheet.

- Explore data set

Browse and Edit Data Buttons

Delete data in Edit view

## Keeping Track of your STATA session

- log file

File > Log > Begin (smcl)

File > Log > Suspense

File > Log > Close

File > Log > View and find a location

- print

In smcl file, edit > copy and paste to the document file

- do file

do edit in command line

do file is the sequence of command lines (record every command in your analysis in the do file. So, you can replicate the result in the future.)

## Getting Help in Stata

In the command line, type help command\_name, e.g. **help summarize**

In the result you will find syntax, option, detail, and example.

Or if you only know what you want to do use **search keyword**

If you remember the command\_name, type **db command\_name**, it will give the command box.

Google search

UCLA Stata Website

# Exploratory Data Analysis in Stata

Key commands

## - **describe**

Getting a lot of information: variable name, variable label, type

Break command

## - **codebook**

A lot more information: type, range, missing data

Discrete

Continuous – numeric (double)

codebook var1 var2 var3

## - **summarize**

Summarize

Dropdown > After choosing the variable and options > Submit and OK

More detail, e.g. median, go back to dropdown: Standard statistics, median (percentile)

Summarize, by group

Summarize if group == 1

- **tabulate** (categorical variables)

Tabulate sex

Dropdown

Tabulate var1 var2

- **list**

All information in the dataset

List var

List var in 1/10

## Create new variables in Stata

Key commands

### - **generate**

Dropdown – create the new variables

Varsq = var\*var

Check the result

Missing data – if the original variable has missing data

### - **replace**

Generate categorical data

Gen cat = .

Replace cat = 1 if var < 40

Replace cat = 2 if var >= 40 & var <50

Replace cat = 3 if var >= 50 & var <60

Replace cat = 4 if var >= 60 & var <70

### **- label variable**

Label variable cat “XXX”

### **- label define**

Label define catlabel 1 “<40” 2 “40-49” 3 “50-59” 4 “60-70”

### **- label values**

Label values cat catlabel

## Graphing in Stata

Key concepts

- Boxplot
- Histogram
- Scatter plot
- Graph editor (You can modify your graph.)
- Saving graphs (both gph and png/jpg/pdf)

## 2. กราฟสำหรับตัวแปรต่อเนื่องตัวเดียว (One Continuous Variables)

---

- การสร้างภาพข้อมูลสำหรับตัวแปรต่อเนื่อง เป็นองค์ประกอบสำคัญสำหรับการสำรวจข้อมูล (Exploratory Data Analysis)
- ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Functions: PDF)
- กราฟ เช่น ฮิสโตแกรม (Histogram), กราฟความหนาแน่น (Density Plot), กราฟควอไทล์ (Quartile) เพื่อตรวจสอบลักษณะของการกระจายตัวของข้อมูล
- แนวคิดสถิติที่สำคัญ ได้แก่
  - ค่าเฉลี่ย (Mean), มัธยฐาน (Median), และ ฐานนิยม (Mode)
  - ความเบี้ยว (Skewness) ซึ่งบ่งบอกว่าข้อมูลกระจุกตัวไปทางด้านใด
  - ความแหลมของการกระจาย (Kurtosis) ซึ่งบ่งบอกว่าข้อมูลมีทางหนามากน้อยเพียงใด

# Histogram

จัดกลุ่มค่าตัวเลขให้อยู่ในช่วงต่างๆ (เรียกว่า bin) ทำให้สามารถเข้าใจการกระจายตัวของตัวแปรต่อเนื่องได้ง่าย แกน X แสดงค่าตัวเลข (เช่น ค่าจ้าง) ส่วนแกน Y มักแสดง ความหนาแน่น (Density)

## ข้อดี

1. เข้าใจง่าย
2. สรุปข้อมูลได้ดี
3. สามารถปรับแต่งได้โดยการเปลี่ยนขนาด bin หรือจำนวน bin

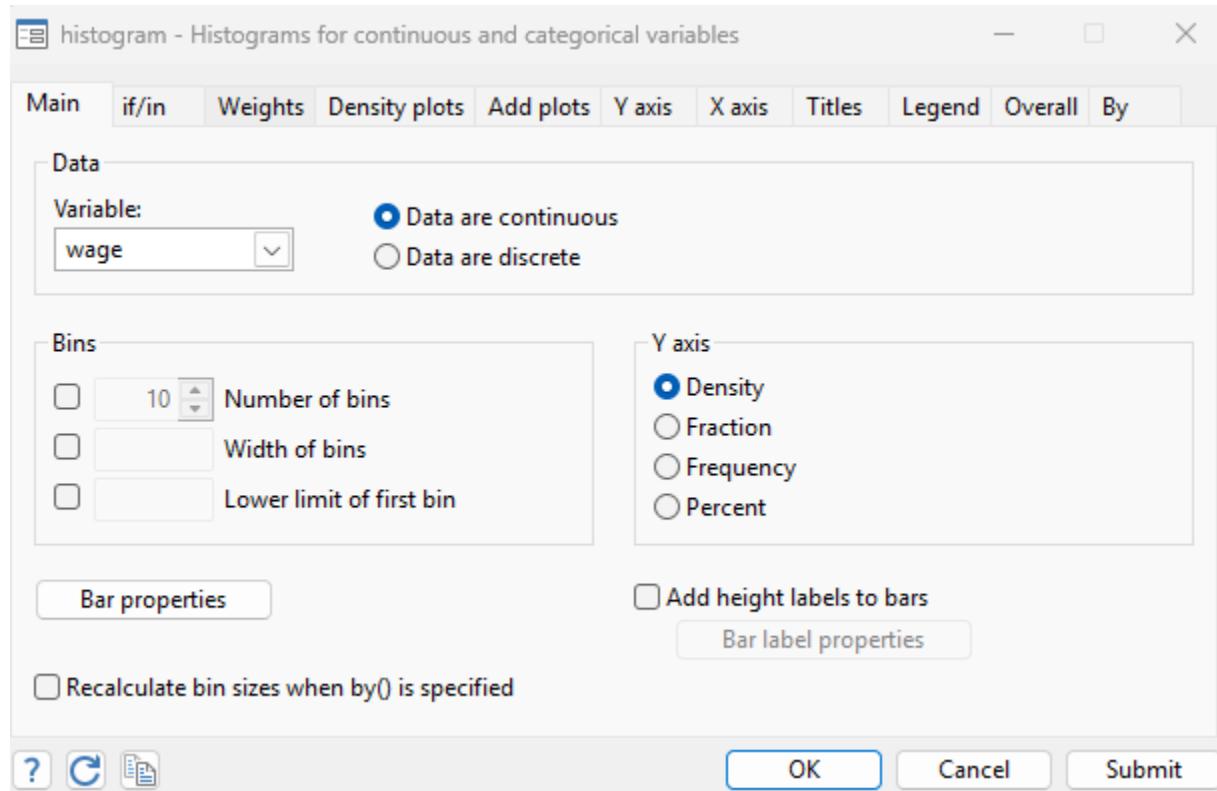
## ข้อเสีย

1. เปรียบเทียบหลาย histogram ได้ยาก
2. การกำหนดขนาดของ bin เป็นเรื่องอัตโนมัติ
3. ค่าตั้งต้นในโปรแกรมมักกำหนดแกน Y เป็นค่าความถี่ซึ่งอาจจะเข้าใจยาก
4. ไม่สามารถดูค่าตัวเลขเฉพาะจุดได้โดยตรง

ใช้ข้อมูลในโปรแกรม Stata เป็นข้อมูลแรงงาน เราสนใจค่าจ้าง (**wage**)

**sysuse nlsw88, clear**

เราสามารถเลือก Graphich > Histogram แล้วเลือกตัวแปร wage

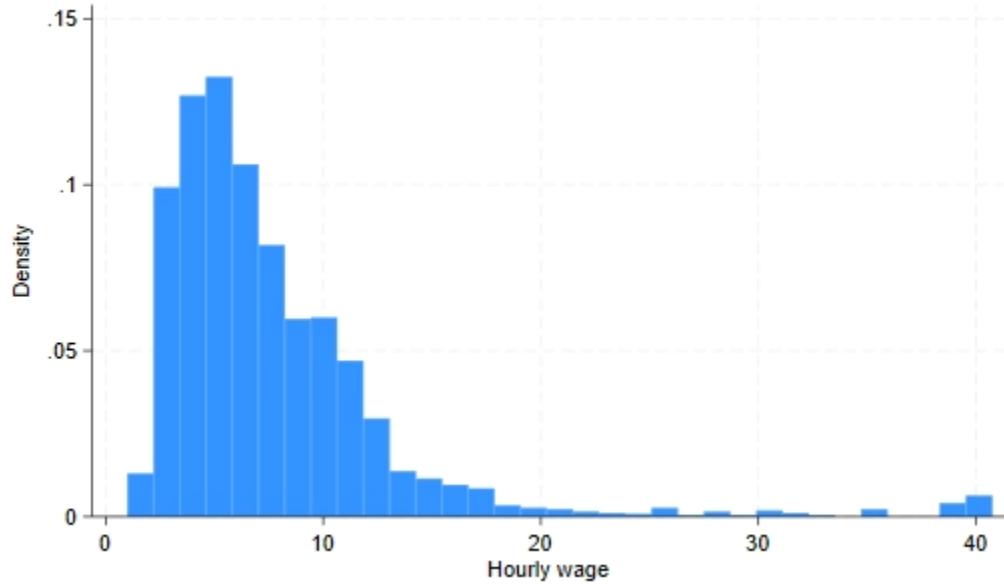


แล้วเลือก Submit

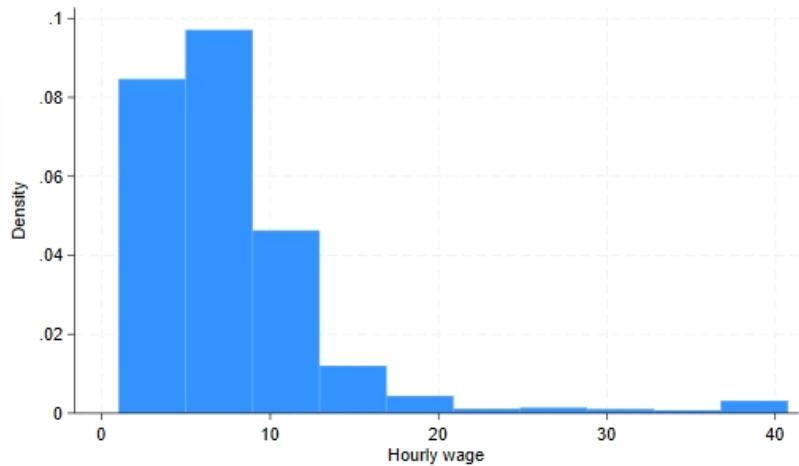
ໃນ Command Window ຈະໄດ້

. histogram wage

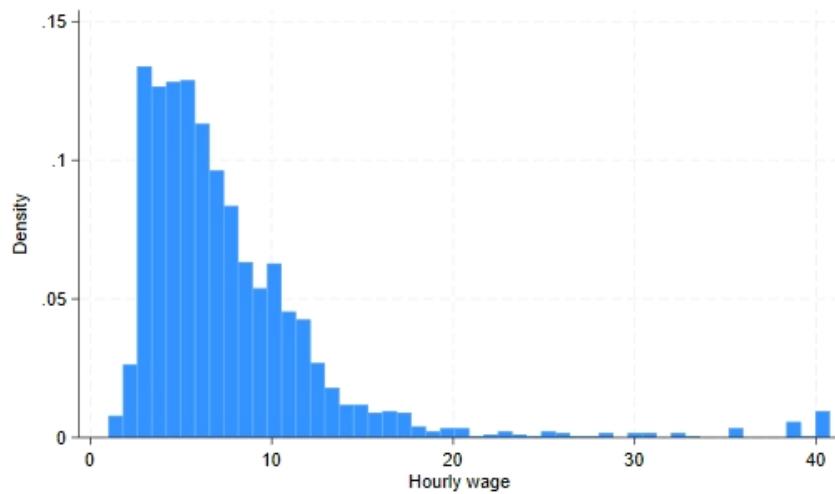
(**bin=33, start=1.0049518, width=1.2042921**)



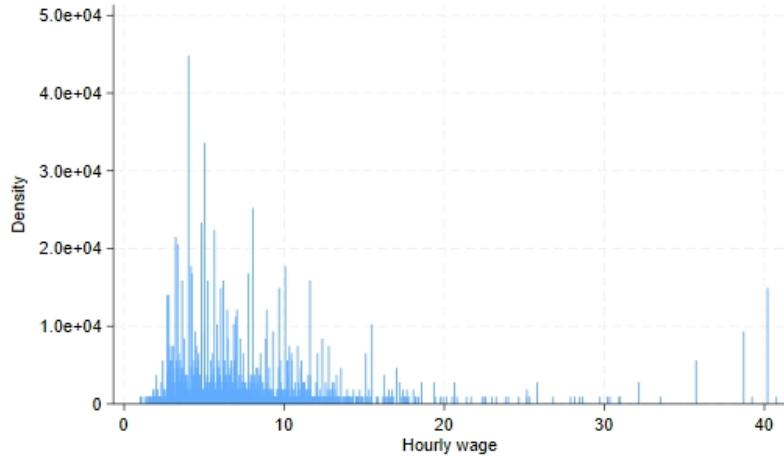
**histogram wage, bin(10)**



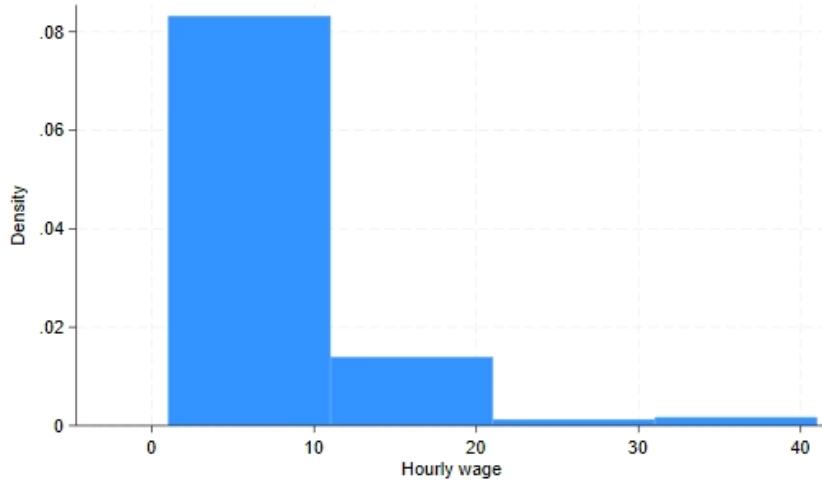
**histogram wage, bin(50)**



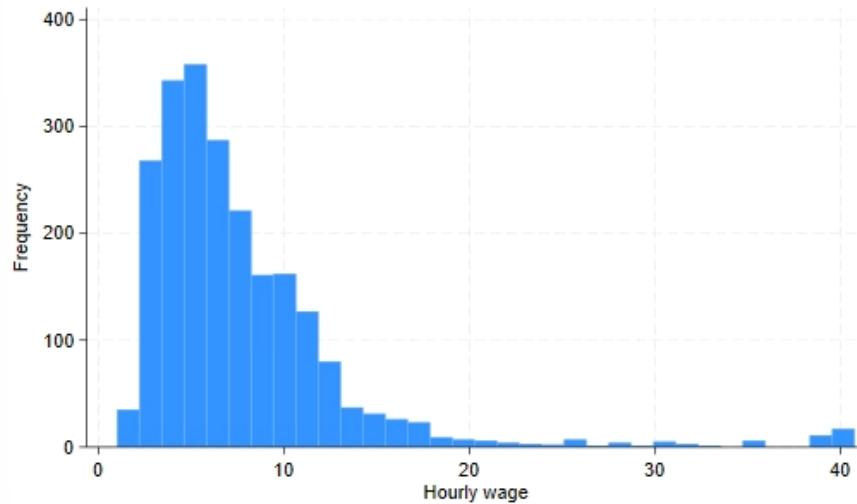
**histogram wage, discrete**



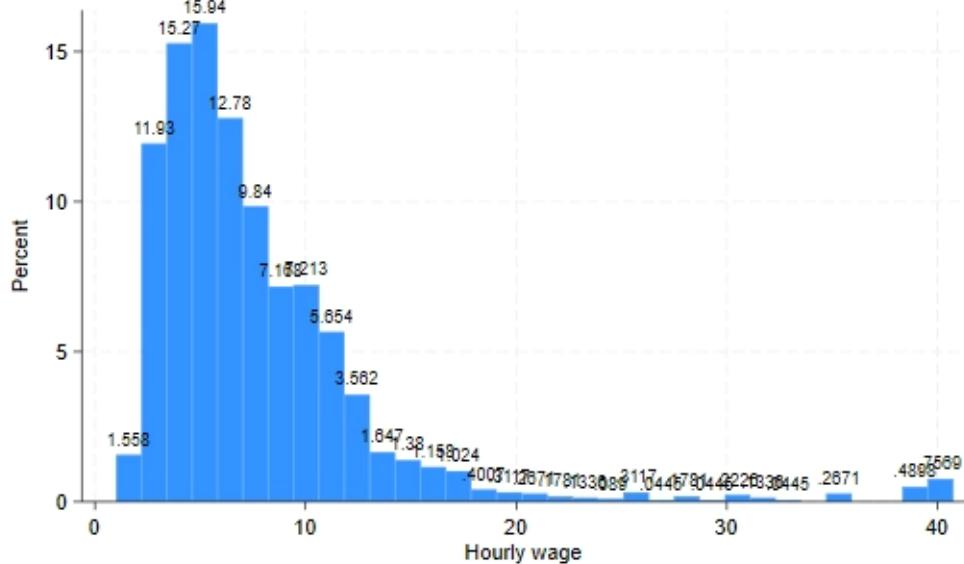
**histogram wage, width(10)**



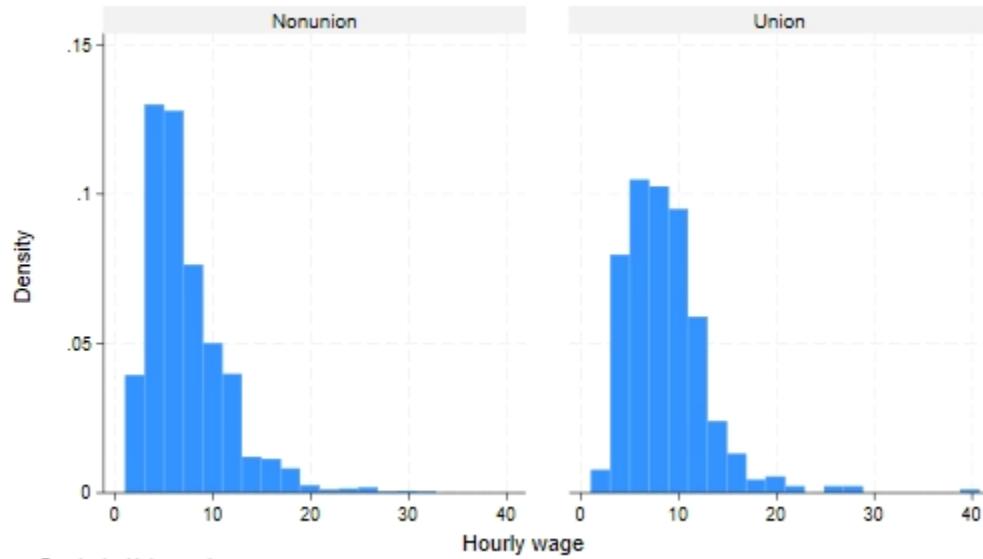
**histogram wage, freq**



**histogram wage, percent addlabels**

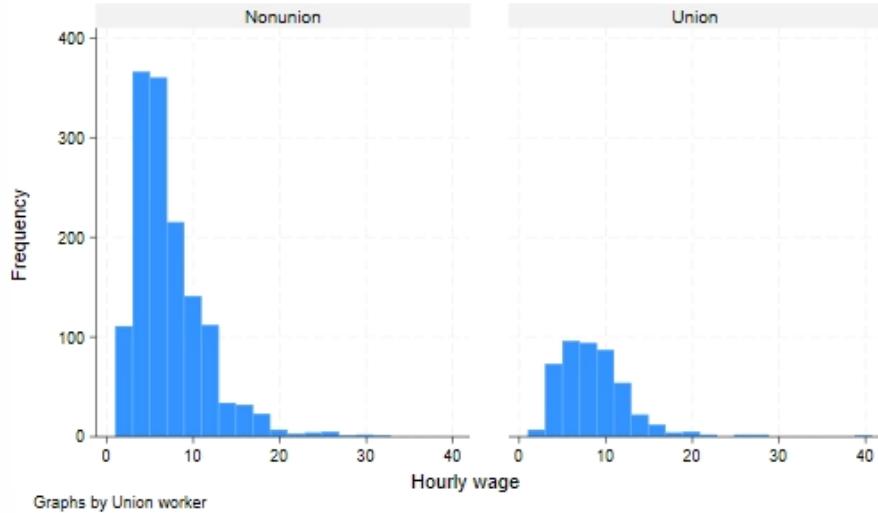


histogram wage, bin(20) by(union)



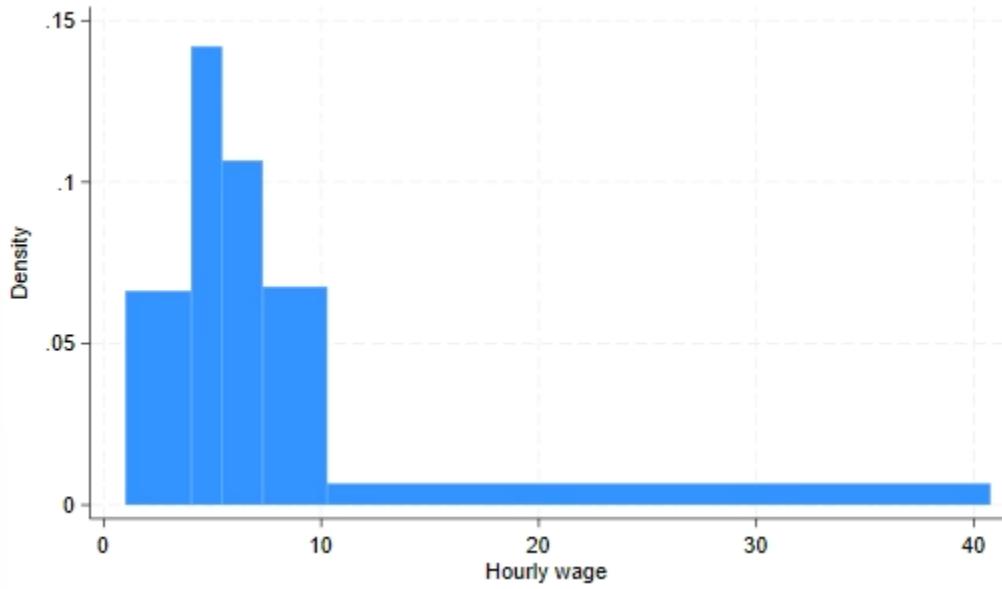
Graphs by Union worker

```
histogram wage, freq bin(20) by(union)
```



\* Equal probability histogram

```
ssc install eqphistogram  
eqphistogram wage, bin(5)
```



### ฮิสโตแกรมแบบความน่าจะเป็นเท่ากัน (Equal Probability Histogram)

- แต่ละ bin มีข้อมูลปริมาณเท่ากัน เช่น  $5 \text{ bin} = \text{bin} \text{ ละ } 20\%$
- ข้อดี: เข้าใจง่าย แสดงช่วงเบอร์เซ็นไทล์ชัดเจน
- ข้อเสีย: มองไม่เห็นรายละเอียดภายในช่วงกว้างของ bin

## Density Plot

Kernel Density Estimation (KDE) หรือ Density Plot ใช้เทคนิค Kernel Smoothing ทำให้เกิดเส้น

# โคงที่ต่อเนื่อง

## ข้อดี

1. เป็นทางเลือกที่رابเรียบกว่าไฮสโตแกรมในการแสดงการกระจายตัวของข้อมูลต่อเนื่อง
2. ช่วยให้เห็น จุดสูงสุด (Peak) และจุดต่ำสุด (Trough)
3. ลดอคติมากกว่าไฮสโตแกรม
4. สวยงามกว่า

## ข้อเสีย

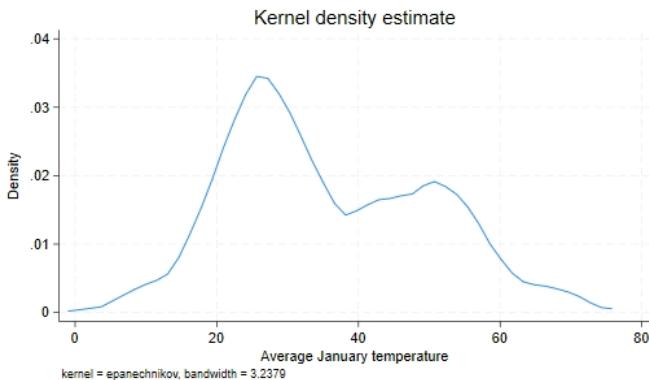
1. อาจแสดงค่าที่ไม่มีอยู่จริง (บริเวณปลายของการกระจาย) หรือสร้างค่าที่เป็นไปไม่ได้ (เช่นค่าลับในข้อมูลที่มีแต่ค่าบวก)
2. ต้องเลือก Bandwidth (ระดับการทำให้กราฟเรียบ) อย่างระมัดระวัง
3. ความซับซ้อน ต้องเลือก Bandwidth และ Kernel (วิธีการกำหนดค่าน้ำหนัก)
  - a. Bandwidth ใหญ่กราฟราบเรียบกว่าแต่เสียรายละเอียด (เล็กกราฟจะละเอียดแต่ชรุขระ)

- b. ควรทดลองค่าต่าง ๆ เพื่อหาความสมดุล
- c. การเลือก Kernel ที่ใช้บ่อยสุดคือ Gaussian Kernel (ราบรื่นที่สุด) กับ Rectangular Kernel (ให้น้ำหนักทุกค่าภายใน Bandwidth เท่ากัน)

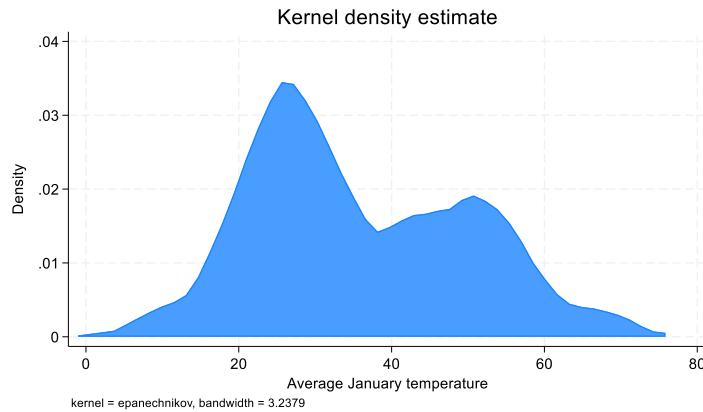
```
sysuse citytemp.dta, clear
```

Dropdown: Graphics > Smoothing and Densities > Kernel density estimation

```
kdensity tempjan
```

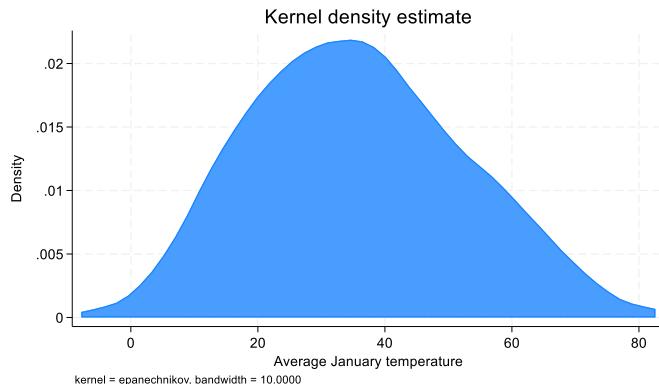


```
kdensity tempjan, recast(area)
```

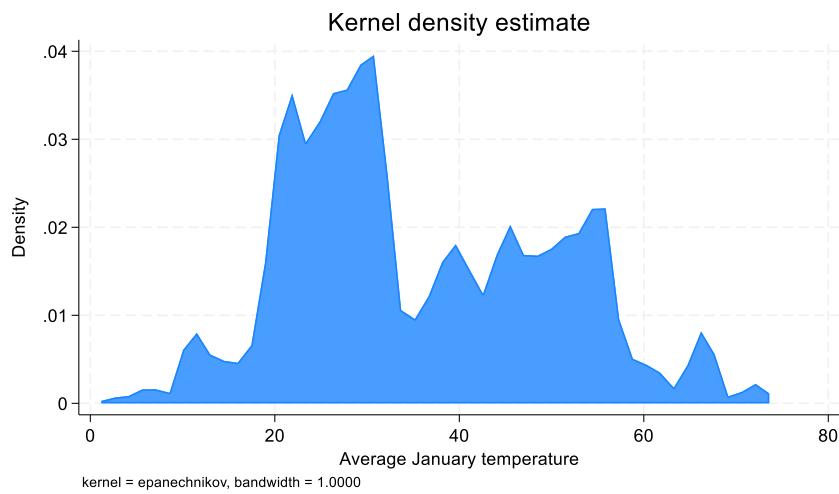


กราฟที่มี bandwidth ต่างกัน

**kdensity tempjan, recast(area) bw(10)**

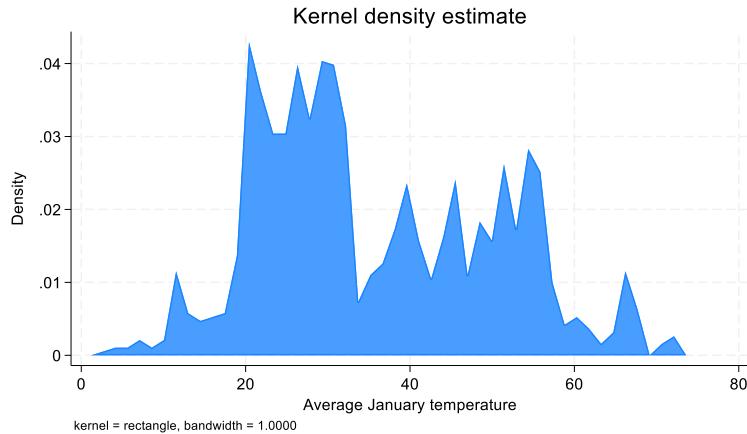


**kdensity tempjan, recast(area) bw(1)**

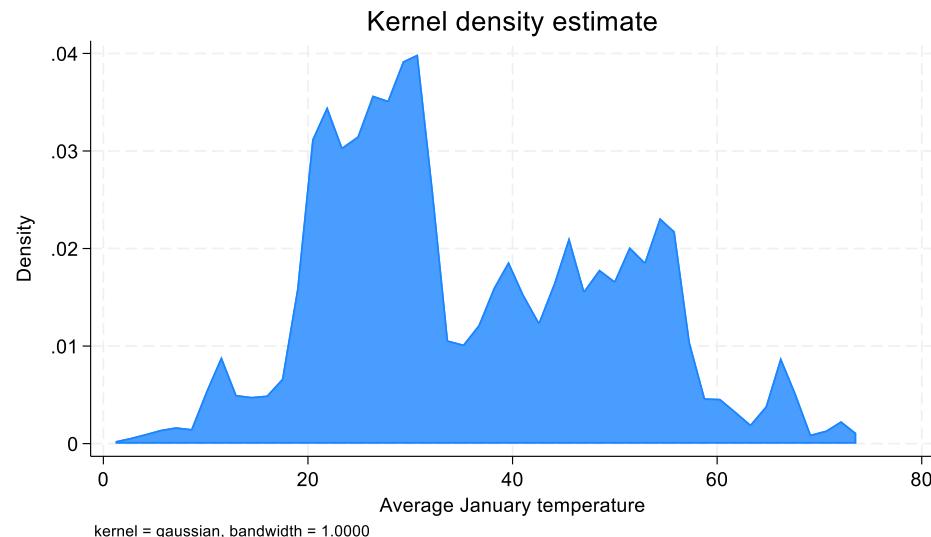


กราฟที่มี kernel ที่ต่างกัน

**kdensity tempjan, recast(area) bw(1) kernel(rectangle)**



**kdensity tempjan, recast(area) bw(1) kernel(gaussian)**



## Multiple Density Plot

เมื่อเปรียบเทียบการกระจายของข้อมูลหลายชุด วิธีการนำเสนอ มีความสำคัญต่อความชัดเจน หากแสดงฮิสโตแกรมหรือกราฟความหนาแน่นแยกกัน การเปรียบเทียบจะทำได้ยาก ควรใช้วิธีที่เหมาะสมกว่า

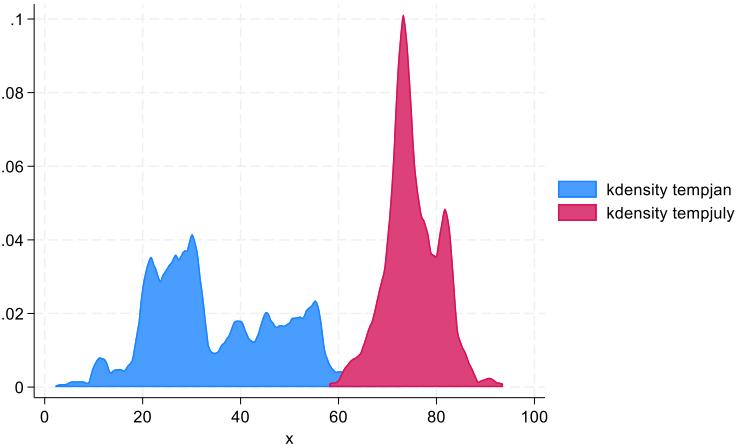
- กราฟความหนาแน่นที่ซ้อนทับกัน (Overlapping density plots)
- กราฟแบบกระจกเงา (Mirror Plots)

ข้อแนะนำ

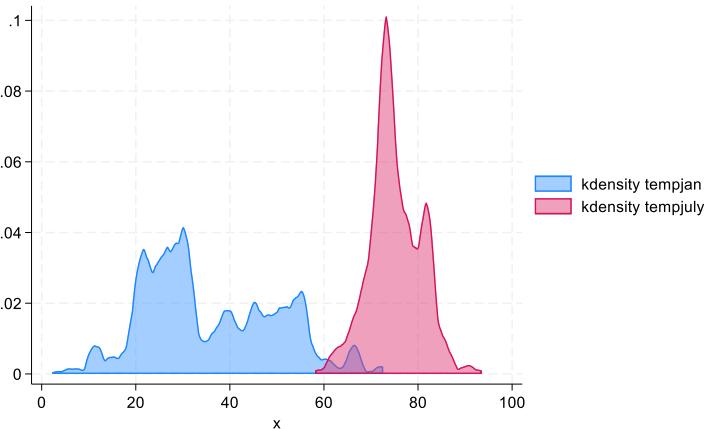
- ใช้ได้ดีถ้าการกระจายของข้อมูลแยกออกจากกันชัดเจน
- ใช้สีเพื่อช่วยแยกข้อมูล แต่ไม่ควรใช้มากเกินไป
- ทางเลือก: ใช้เส้นประหรือเส้นจุดแทนสี

## ตัวอย่าง Multiple Density Plot

```
twoway kdensity tempjan, recast(area) bw(1) || kdensity tempjuly,  
recast(area) bw(1)
```



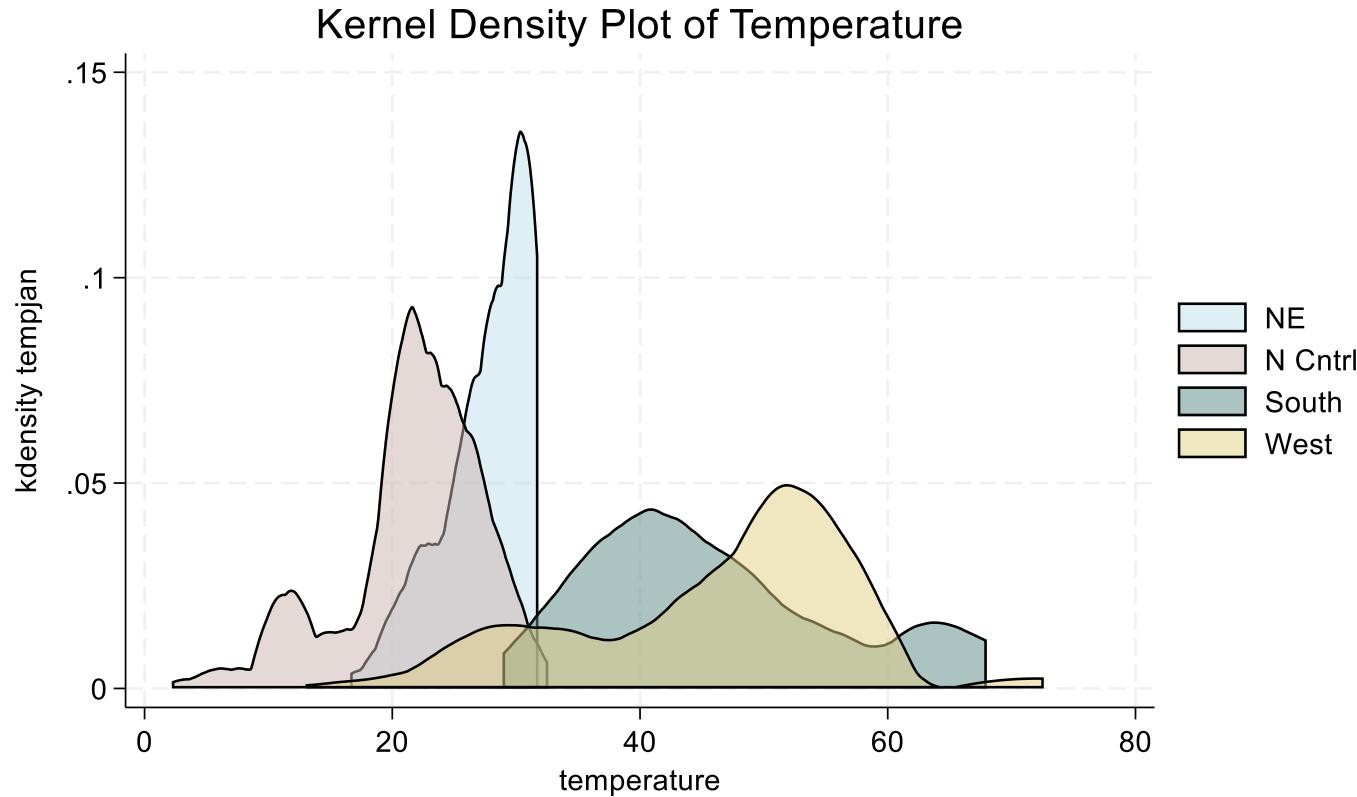
```
twoway kdensity tempjan, recast(area) bw(1) fc(%50) || kdensity  
tempjuly, recast(area) bw(1) fc(%50)
```



```

twoway ///
(kdensity tempjan if region == 1, col(black) fc(ltblue%50) recast(area)) ///
(kdensity tempjan if region == 2, col(black) fc(erose%50) recast(area)) ///
(kdensity tempjan if region == 3, col(black) fc(emerald%50) recast(area)) ///
(kdensity tempjan if region == 4, col(black) fc(sand%50) recast(area)), ///
title("Kernel Density Plot of Temperature") ///
xtitle("temperature") ///
legend(order(1 "NE" 2 "N Cntrl" 3 "South" 4 "West"))

```



## Ridgeline Plot (Joy Plots)

Ridgeline Plots (หรือที่เรียกว่า Joy Plots) เป็นวิธีแสดง Kernel Density Plots หลายชุด ในรูปเดียว หลักการทำงานคือการวาง Density Plots หลายตัวบนแกน X แยกกัน แต่ยังคงอยู่ในรูปเดียว กัน ช่วยให้สามารถเปรียบเทียบการกระจายของข้อมูลหลายชุดได้ง่ายขึ้น

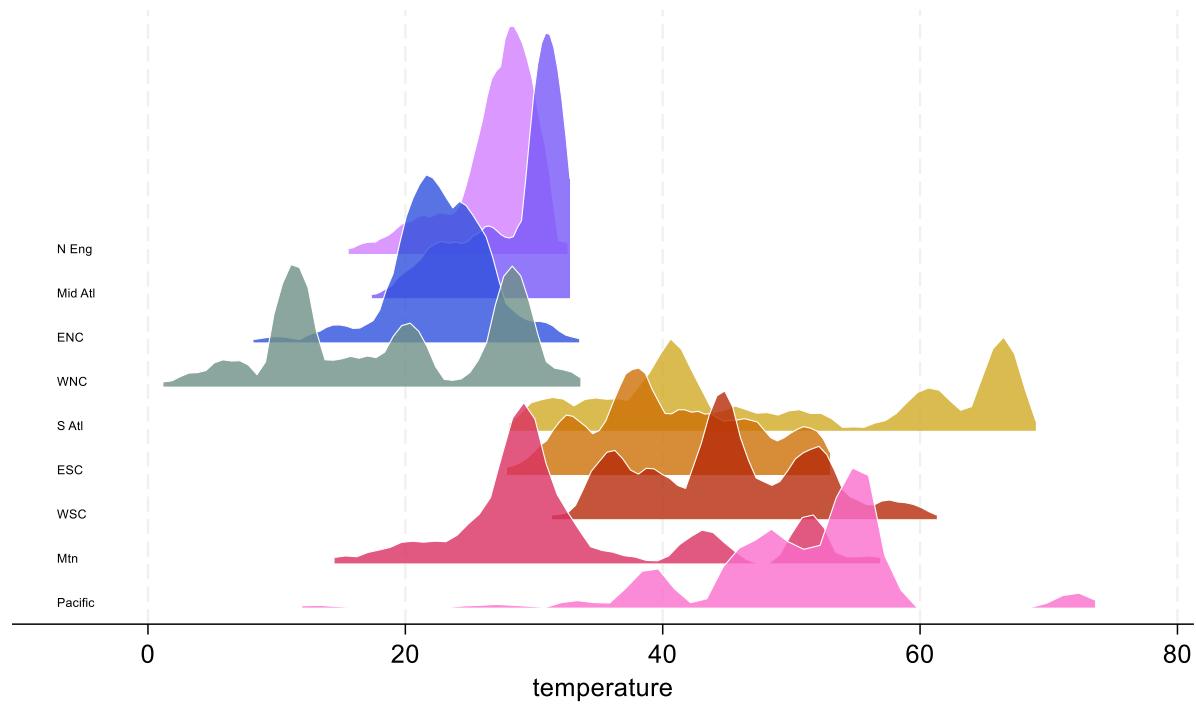
### ข้อดี

- เหมาะสมสำหรับเปรียบเทียบการกระจายหลายชุดข้อมูล (ใช้ได้เมื่อมีหลายกลุ่มแต่ไม่มากเกินไป)
- แสดงแนวโน้มและความแตกต่างของกลุ่มข้อมูลได้ดี เช่น การเปลี่ยนแปลงตามเวลา
- กราฟดูสะอาดและเข้าใจง่ายกว่าฮิสโตแกรมที่ซ้อนทับกัน

### ข้อเสีย

- ถ้าไม่มีการจัดเรียงข้อมูลที่ดี กราฟอาจดูยุ่งเหยิงได้
- ต้องปรับขนาดแกน X และ Y อย่างระมัดระวัง และเลือก Bandwidth ที่เหมาะสม
- ถ้ามีการซ้อนทับของกลุ่มข้อมูลมากเกินไป อาจทำให้กราฟอ่านยาก

```
ssc install joyplot  
ssc install palettes  
ssc install colrspace  
joyplot tempjan, by(division) bwidth(1)  
xtitle("temperature")
```



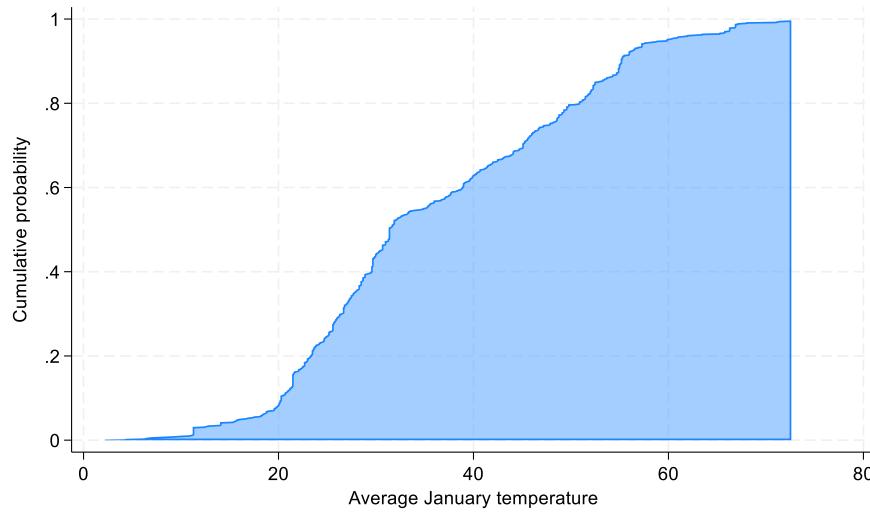
## Cumulative Density Plot

Cumulative Density Plots (CDF หรือ Empirical CDF) เป็นกราฟที่แสดงเปอร์เซ็นต์สะสมของข้อมูลตามค่าของตัวแปร ซึ่งต่างจาก density plot ที่แสดงสัดส่วนของข้อมูลในแต่ละค่า

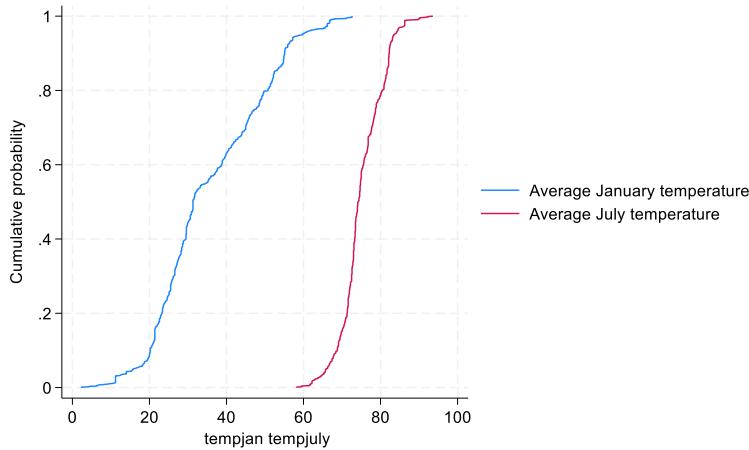
CDF แสดงการสะสมของข้อมูลไปเรื่อย ๆ และจะจบที่ค่า 1 เสมอ

- จุดที่กราฟ ชันที่สุด แสดงว่า มีข้อมูลกระจุกตัวมากที่สุด
- ช่วงราบที่ตันและปลาย แสดงว่า ข้อมูลในช่วงนั้นมีจำนวนน้อย (ช่วยในการตรวจจับ Outliers ช่วงราบที่ยาว)
- สามารถอ่านค่าสถิติที่สำคัญ เช่น เปอร์เซ็นไทล์ มัธยฐาน ค่าต่ำสุด/สูงสุด ได้โดยตรงจากการ

```
ssc install distplot  
distplot tempjan, recast(area) fc(%50)
```



```
distplot tempjan tempjuly
```



## Spike Plot

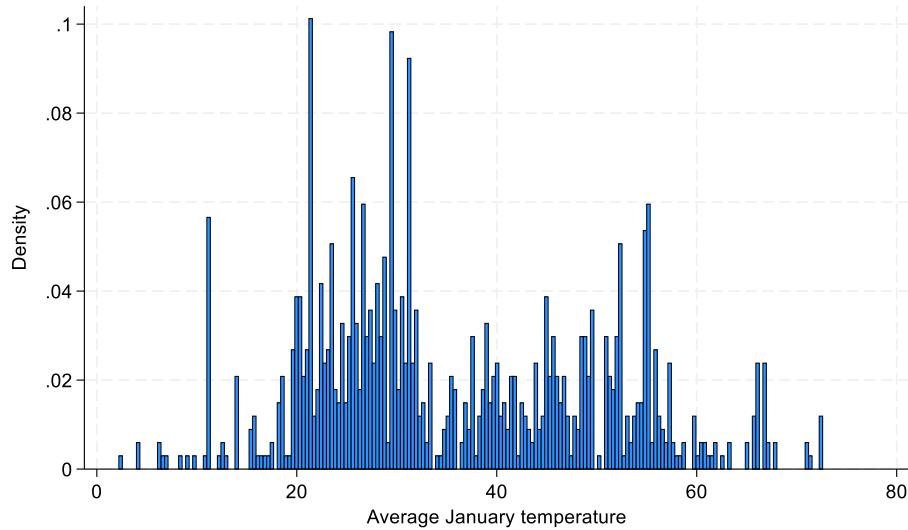
Spike Plots คือ ฮิสโตแกรมที่มีจำนวน bin สูงมาก (100–500 bin)

ใช้เพื่อตรวจสอบรายละเอียดที่เล็กมากในชุดข้อมูล

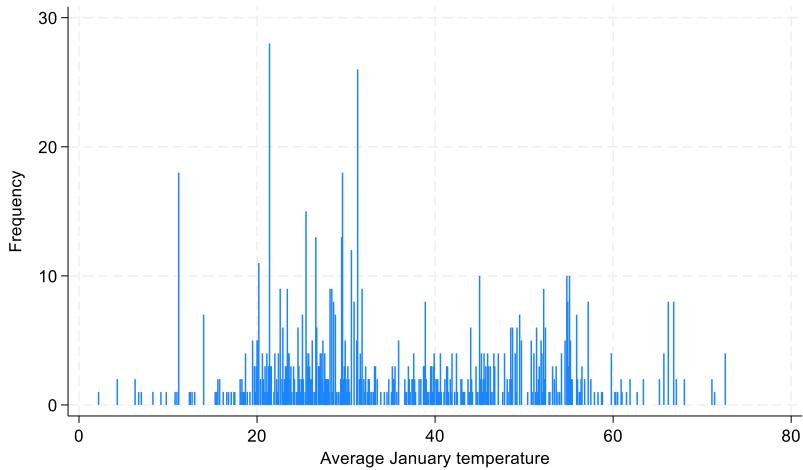
แนวคิดหลักและการใช้งานของ Spike Plots:

- ฮิสโตแกรมทั่วไป ( เช่น 30 bin ) ให้มุ่งมองกว้างของการกระจายข้อมูล
- Spike Plot ( เช่น 200+ bin ) ใช้ตรวจสอบรูปแบบที่ละเอียดมากขึ้น
- ตัวอย่าง: หากฮิสโตแกรมของค่าจ้างแสดงค่าซ่อง 30–40, Spike Plot อาจเผยให้เห็นว่ามีค่าจ้างที่พบบ่อยที่ 38.8 และ 40.8 ซึ่งอาจเกี่ยวข้องกับ สัญญาจ้างแบบสหภาพแรงงาน

**histogram tempjan, bin(200) lc(black) lw(vvthin)**

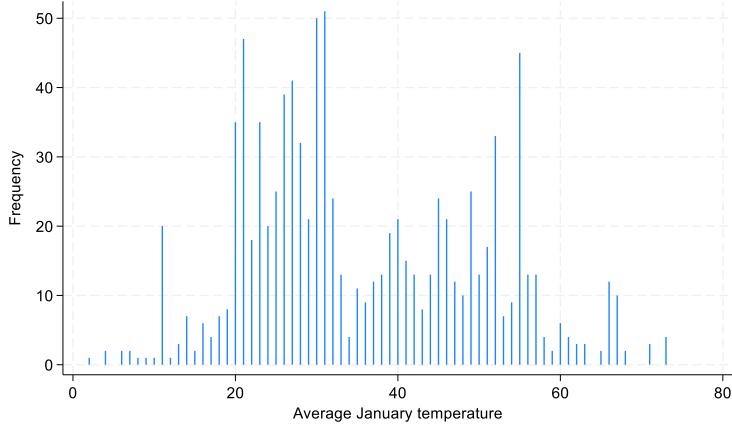


**spikeplot tempjan**



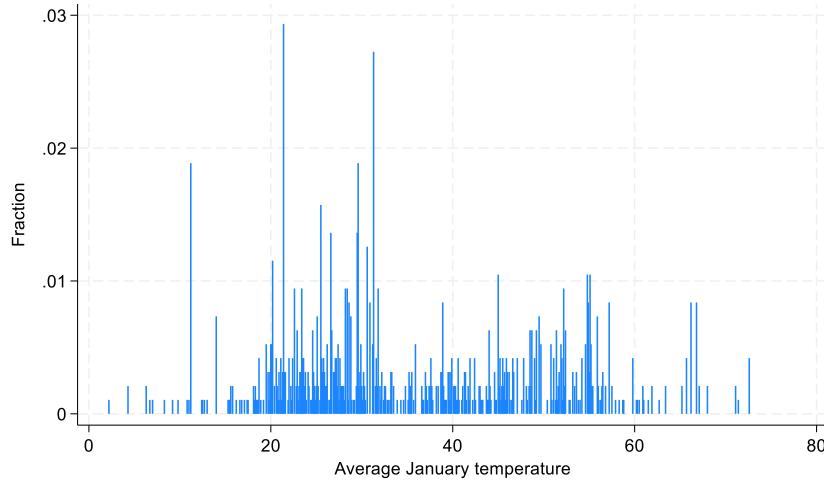
Spike plot with different bandwidth

**spikeplot tempjan, round(1)**



Spike plot with fraction on Y-axis

**spikeplot tempjan, fraction**



## Box Plot

Box Plot (หรือ Box-and-Whisker Plot) เป็น เครื่องมือวิเคราะห์การกระจายข้อมูล ที่สรุปข้อมูลโดยใช้ 5 ค่าทางสถิติสำคัญ ได้แก่:

1. ค่าต่ำสุด (ยกเว้น Outlier)
2. ควอไทล์ที่ 1 (Q1, 25th percentile)
3. มัธยฐาน (Q2, 50th percentile)
4. ควอไทล์ที่ 3 (Q3, 75th percentile)
5. ค่าสูงสุด (ยกเว้น Outlier)

Box Plot มี เส้น Whiskers เพื่อแสดงช่วงค่าต่ำสุด/สูงสุด และจุด Outliers ซึ่งมักถูกกำหนดเป็น 1.5 เท่า ของช่วงควอไทล์ (IQR)

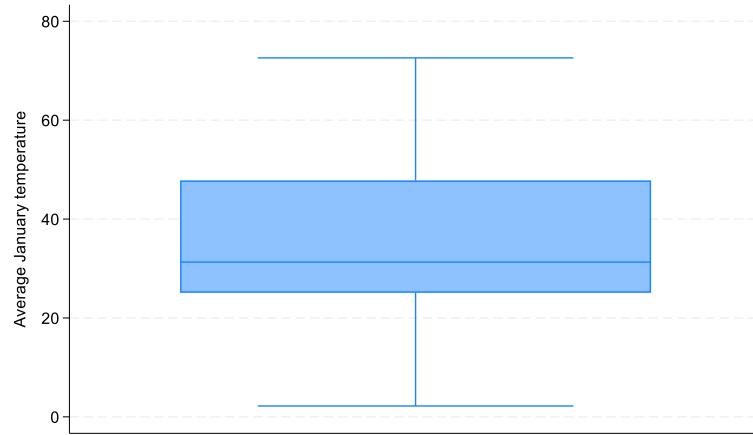
## ข้อดี

- เหมาะกับการเปรียบเทียบหลายกลุ่มข้อมูล
- สื่อสารสถิติสำคัญได้ดี
- ช่วยตรวจจับ Outlier ได้ดี
- สามารถใช้ร่วมกับกราฟอื่นๆ

## ข้อเสีย

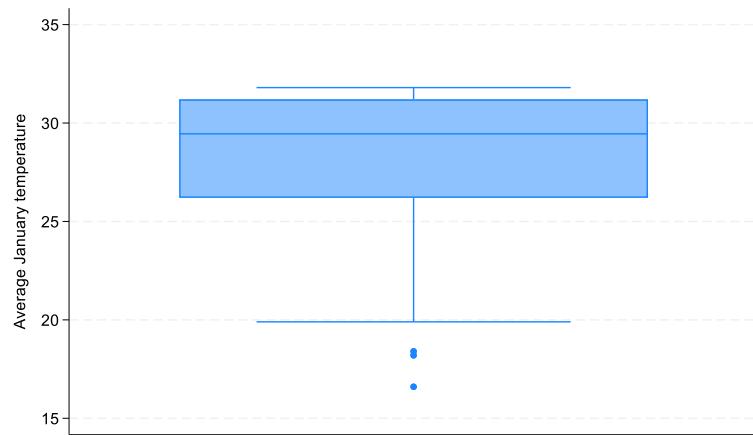
- ไม่แสดงรูปร่างของการกระจายข้อมูลอย่างละเอียด (อาจจะแก้ได้ด้วย Violin Plot)
- ไม่มีค่ามัธยมเลขคณิต (Mean) โดยอัตโนมัติ – อาจทำให้เกิดความสับสนระหว่าง Median กับ Mean
- การกำหนด Outlier อาจแตกต่างกันในแต่ละซอฟต์แวร์

```
graph box tempjan
```



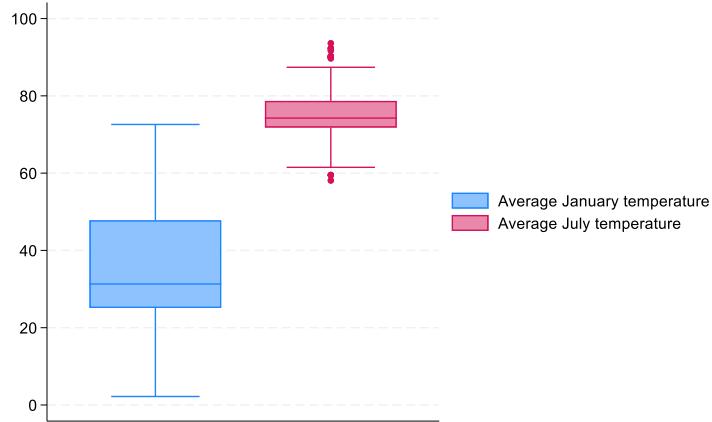
Box plot with outlier

```
graph box tempjan if region == 1
```



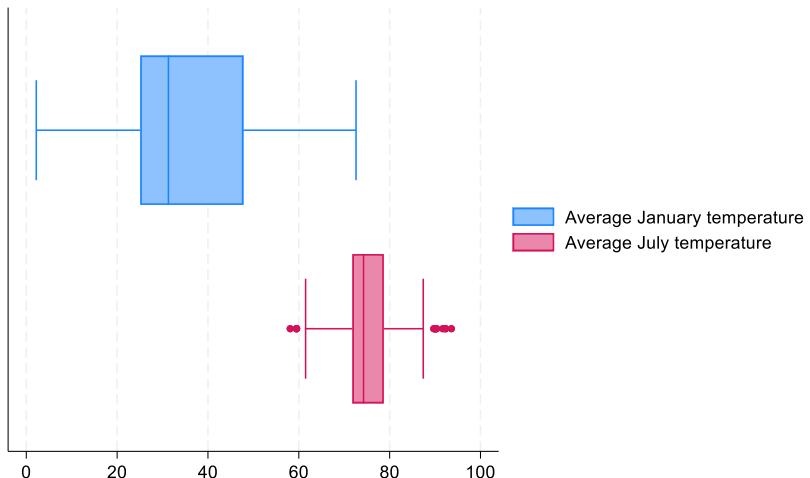
Multiple box plots

**graph box tempjan tempjuly**

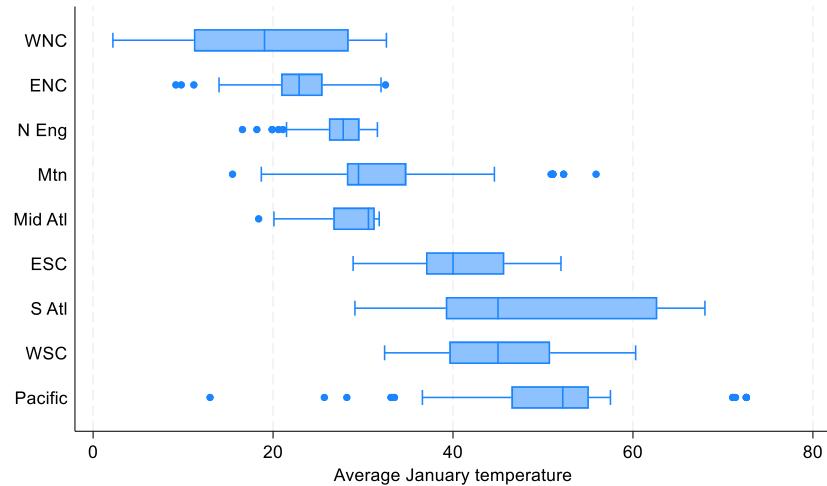


Horizontal box plot

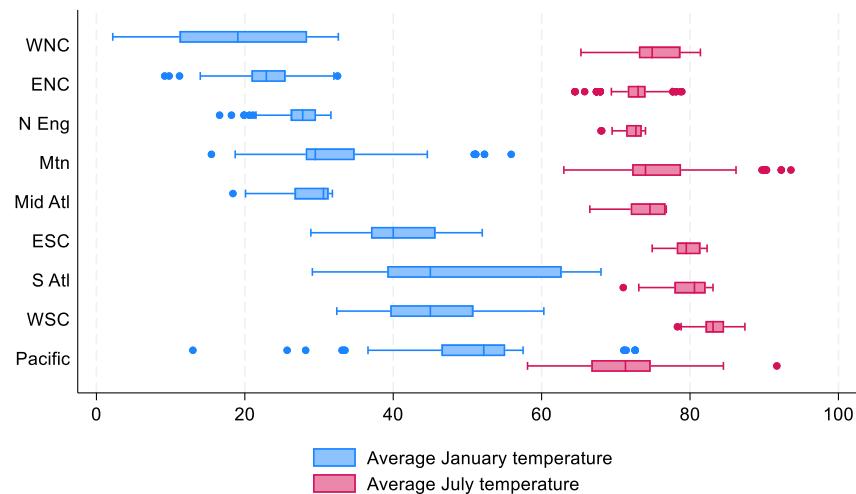
**graph hbox tempjan tempjuly**



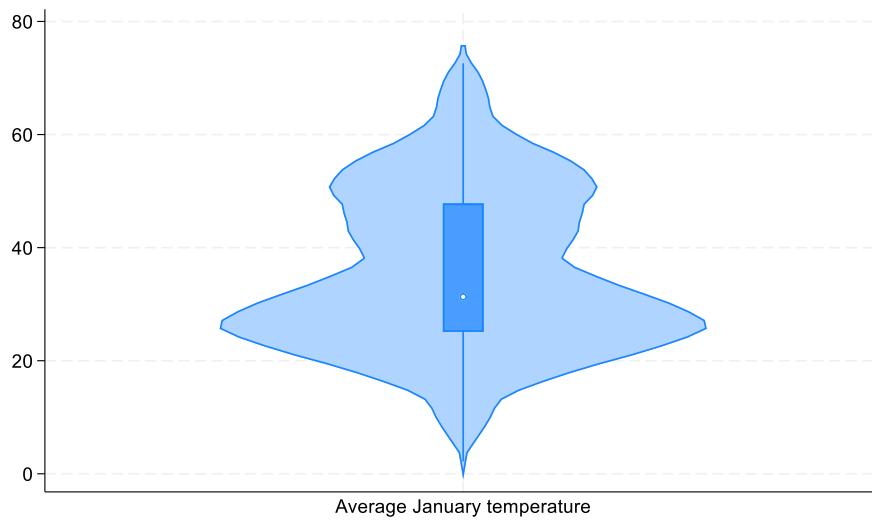
**graph hbox tempjan, over(division, sort(1))**



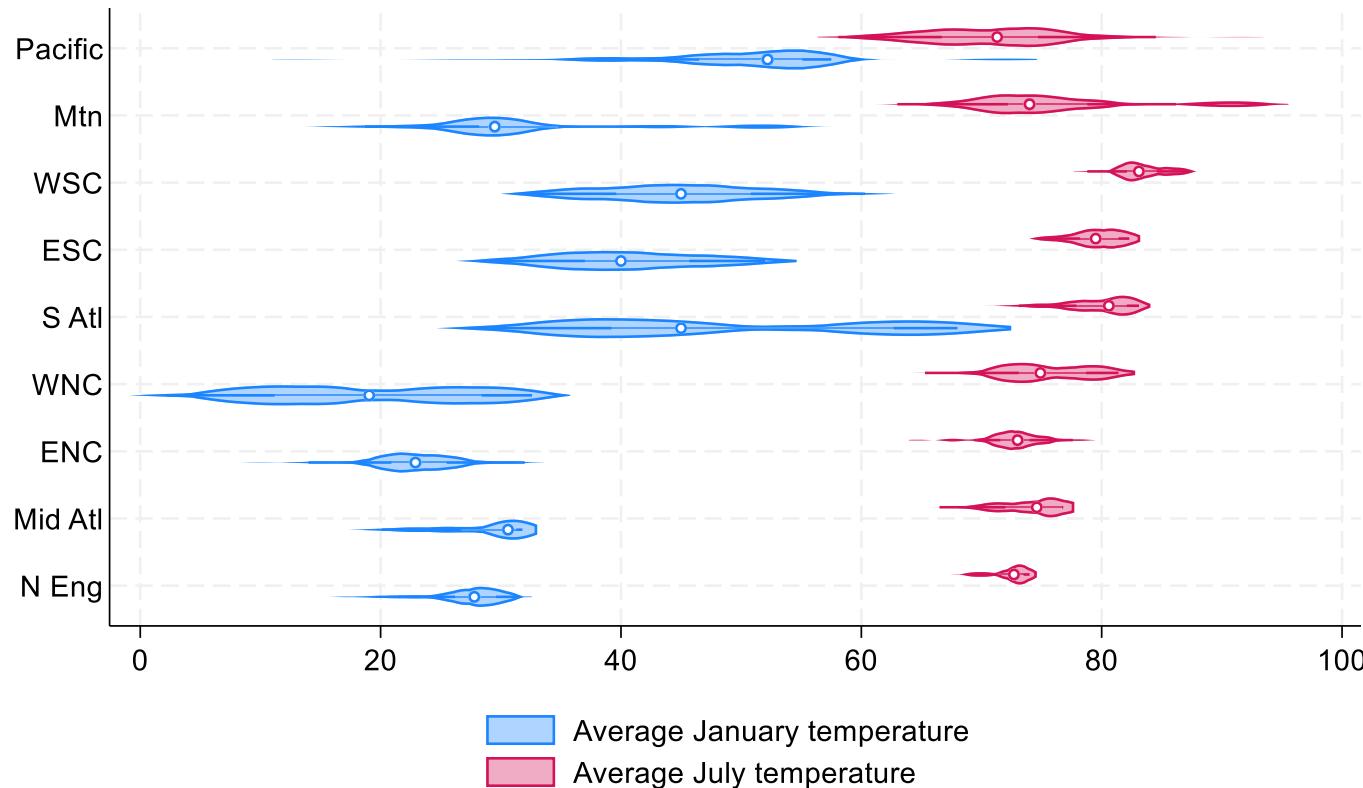
**graph hbox tempjan tempjuly, over(division, sort(1))  
legend(position(6))**



```
ssc install vioplot  
vioplot tempjan
```



## vioplot tempjan tempjuly, over (division) horizontal legend (position (6))



\* Edit y-axis text

## Symmetry Plot

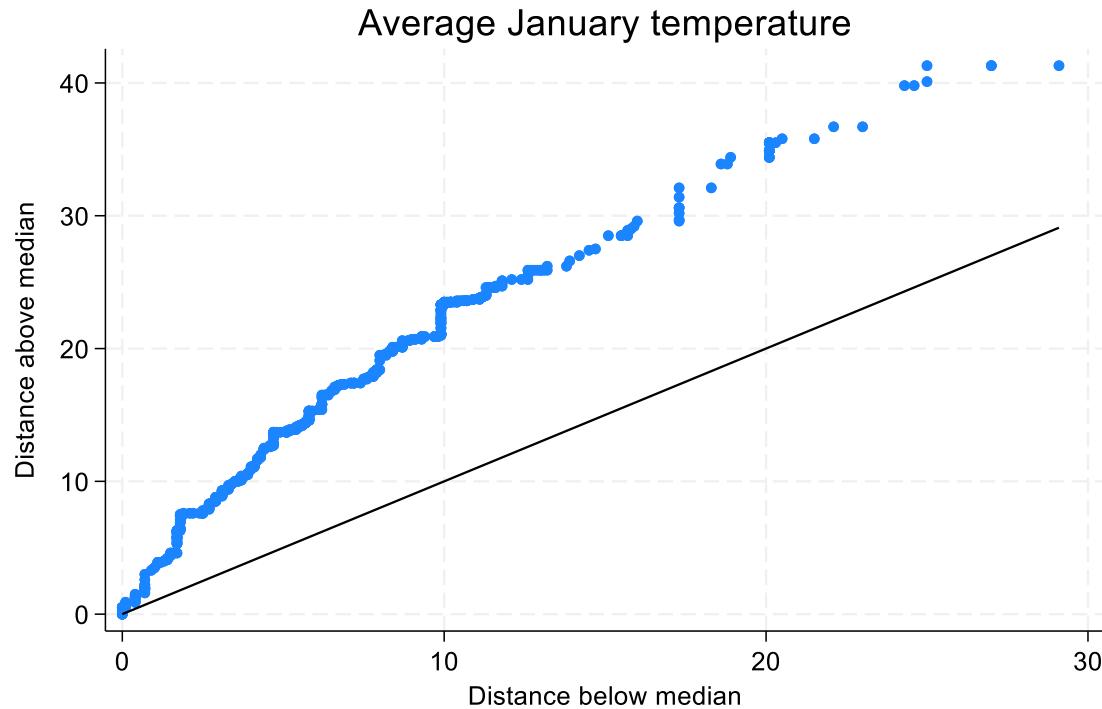
เป็นเทคนิคการสร้างภาพข้อมูลที่ใช้ตรวจสอบว่า การกระจายข้อมูลมีความสมมาตรหรือไม่ รอบค่ามัธยฐาน (Median) กราฟนี้จะเปลี่ยนค่าข้อมูลให้เป็นระยะห่างจากค่ามัธยฐาน และเปรียบเทียบค่าที่อยู่ **ฝั่งซ้าย** ของมัธยฐานกับ **ฝั่งขวา**

หากจุดข้อมูลอยู่บนเส้นอ้างอิง ( $y = x$ ) แสดงว่าการกระจายของข้อมูล สมมาตร

หากจุดเบี่ยงเบนออกไป แสดงว่าข้อมูลมีความเบี้ยว (Skewness)

- ถ้าจุดอยู่เหนือเส้นอ้างอิง  $\rightarrow$  ข้อมูลมี ความเบี้ยวขวา (Right-skewed)
- ถ้าจุดอยู่ต่ำกว่าเส้นอ้างอิง  $\rightarrow$  ข้อมูลมี ความเบี้ยวซ้าย (Left-skewed)

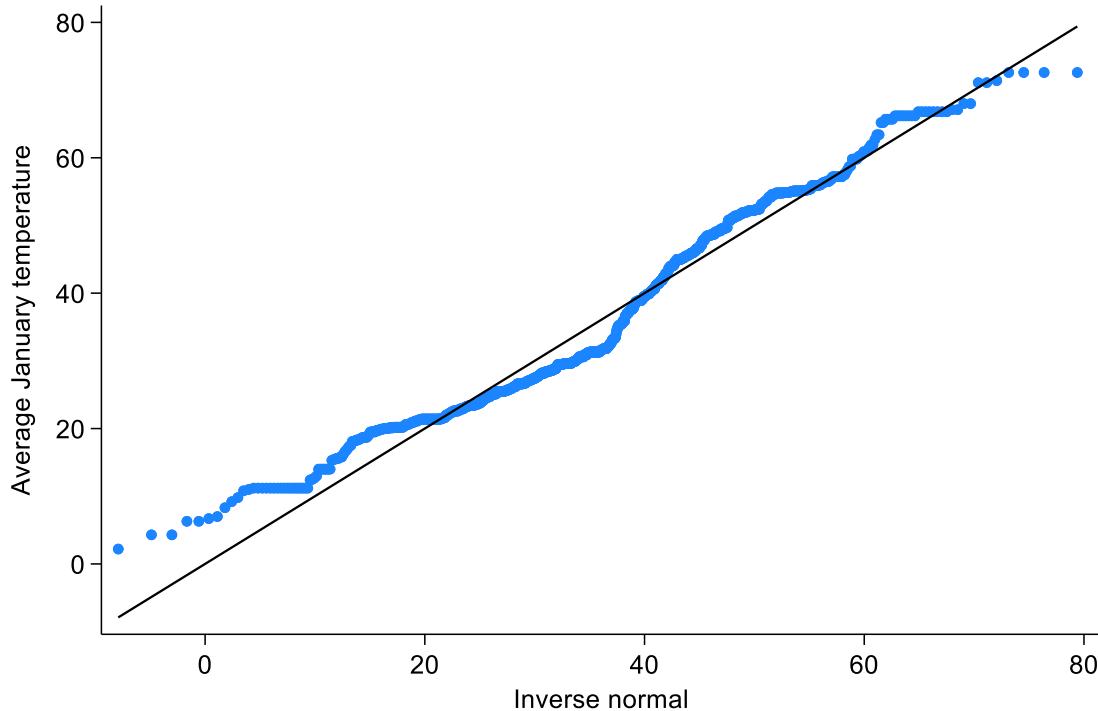
**symplot tempjan**



## Quantile-Normal Plot

เปรียบเทียบว่าการแจกแจงเหมือน Normal distribution ไหม

**qnorm tempjan**



## Quantile-Quantile (QQ) Plot

Quantile-Quantile (QQ) Plot เป็น กราฟกระจาย (Scatter Plot) ที่ใช้เปรียบเทียบ Quantile ของสองชุดข้อมูล เพื่อตรวจสอบว่ามีการกระจายเหมือนกันหรือไม่ หากชุดข้อมูลเหมือนกัน จะจะเรียงตัวบนเส้น 45 องศา ( $y = x$ ) และถ้าข้อมูลมีความสัมพันธ์เชิงเส้นแต่ไม่เหมือนกันทั้งหมด จะจะยังคงเป็นเส้นตรงแต่ขึ้นลงจาก  $y = x$

- แกน X: Quantile ของชุดข้อมูลแรก
- แกน Y: Quantile ของชุดข้อมูลที่สอง
- ถ้าทั้งสองชุดข้อมูลกระจายเหมือนกัน  $\rightarrow$  จะอยู่บนเส้นอ้างอิง  $y = x$

## ข้อดี

- ไม่จำเป็นต้องมีขนาดตัวอย่างเท่ากัน
- ทดสอบการกระจายของข้อมูลได้ง่าย

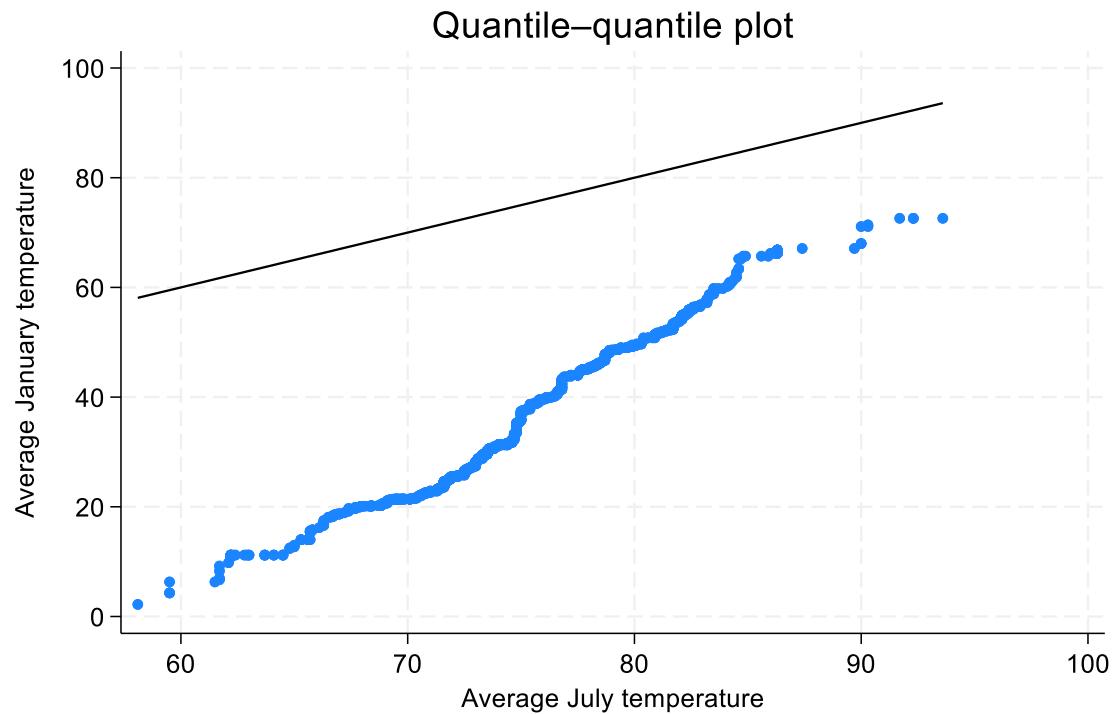
## ข้อเสีย

- เข้าใจยากสำหรับคนทั่วไป และต้องอธิบายให้ดีเมื่อนำเสนอ
- Kernel Density Plot หรือ Histogram อาจทำให้เข้าใจได้ยากกว่า

## ควรใช้เมื่อ

- เปรียบเทียบการกระจายของข้อมูลในช่วงเวลาต่างกัน
- ตรวจสอบว่าการแปลงข้อมูลช่วยให้ข้อมูลเป็น Normal Distribution หรือไม่
- เปรียบเทียบการกระจายของข้อมูลจริงกับทฤษฎี
- ตรวจสอบสมมติฐานทางสถิติในการวิเคราะห์ข้อมูล

## qqplot tempjan tempjuly



## 4. กราฟสำหรับตัวแปรไม่ต่อเนื่องตัวเดียว (One Discrete Variables)

ข้อมูลแบบไม่ต่อเนื่อง (Discrete Data) คือข้อมูลที่ นับได้เป็นจำนวนเต็ม และพบบ่อยในงานวิเคราะห์ ข้อมูล เช่น จำนวนคนในห้อง, วันในสัปดาห์, กลุ่มอายุ, ประเภทอาชีพ

ข้อมูลแบบไม่ต่อเนื่องอาจเป็น ข้อมูลที่มีลำดับ (Ordered) เช่น กลุ่มอายุ หรือ ไม่มีลำดับ (Unordered) เช่น วิธีการเดินทาง (รถไฟ, รถบัส, รถยนต์)

ลักษณะสำคัญของข้อมูลแบบไม่ต่อเนื่อง:

- มีค่าที่เป็นจำนวนจำกัด (เช่น จำนวนลูกในครอบครัว)
- อาจมีลำดับหรือไม่มีลำดับ (เช่น ระดับการศึกษา vs. ประเภทการเดินทาง)
- ไม่ต้องใช้การแบ่งช่วงข้อมูลเหมือนข้อมูลแบบต่อเนื่อง
- นักถูกนำเสนอเป็นข้อมูลประเภทหมวดหมู่ (Categorical Data)

## วิธีการนำเสนอข้อมูล

- ตาราง (One-way Frequency Tables) – อ่านง่ายและยืดหยุ่น สามารถใส่เปอร์เซ็นต์และค่าทางสถิติเพิ่มเติมได้
- แผนภูมิแท่ง (Bar Chart) – ใช้แสดงความถี่หรือเปอร์เซ็นต์ของแต่ละหมวดหมู่
- แผนภูมิวงกลม (Pie Chart) – ใช้แสดงสัดส่วนข้อมูล แต่ไม่เหมาะสมกับข้อมูลที่มีหลายหมวดหมู่
- Dot Plot

# แผนภูมิแท่ง

แผนภูมิแท่ง (Bar Graph หรือ Bar Chart) เป็นเทคนิคการแสดงผลข้อมูลที่ใช้กันมากที่สุดสำหรับข้อมูลแบบไม่ต่อเนื่อง (Discrete Data) และข้อมูลหมวดหมู่ (Categorical Data)

กราฟนี้ใช้แท่งสี่เหลี่ยมที่ความสูงหรือความยาวของแท่งแทนจำนวน/ความถี่ (Frequency) หรือเปอร์เซ็นต์ วดในแนวตั้ง (Column Chart) หรือ แนวนอน (Horizontal Bar Chart)

## ประเภทของแผนภูมิแท่ง:

1. แผนภูมิแท่งพื้นฐาน (Basic Bar Chart) → ใช้แสดงค่าความถี่หรือเปอร์เซ็นต์ของแต่ละหมวดหมู่
2. แผนภูมิแท่งเรียงลำดับ (Sorted Bar Chart) → เรียงลำดับจากมากไปน้อยเพื่อให้เห็นแนวโน้ม
3. แผนภูมิแท่งแนวนอน (Horizontal Bar Chart) → หมายความว่า สำหรับหมวดหมู่ที่มีป้ายชื่อยาว
4. แผนภูมิแท่งแบบซ้อน (Stacked Bar Chart) → ใช้แสดงข้อมูลหมวดหมู่อย่างภายในแต่ละแท่ง
5. แผนภูมิแท่งแบบสะท้อน (Mirrored Bar Chart / Population Pyramid) → ใช้เปรียบเทียบข้อมูลสองกลุ่ม เช่น เพศชาย vs. เพศหญิง

## ข้อดี

- เข้าใจง่าย ใช้ได้กับทุกคน
- เหมาะสำหรับการเปรียบเทียบข้อมูลหลายหมวดหมู่
- ไม่ต้องกำหนดขนาดช่วงข้อมูล (Bin Width)
- สามารถแสดงข้อมูลหลายกลุ่มได้ง่าย

## ข้อเสีย

- อาจทำให้เข้าใจผิดได้ถ้ามีการปรับแกน Y
- ไม่เหมาะสมกับชุดข้อมูลที่มีหมวดหมู่มากเกินไป
- แผนภูมิแห่งซ้อนอาจสับสนได้

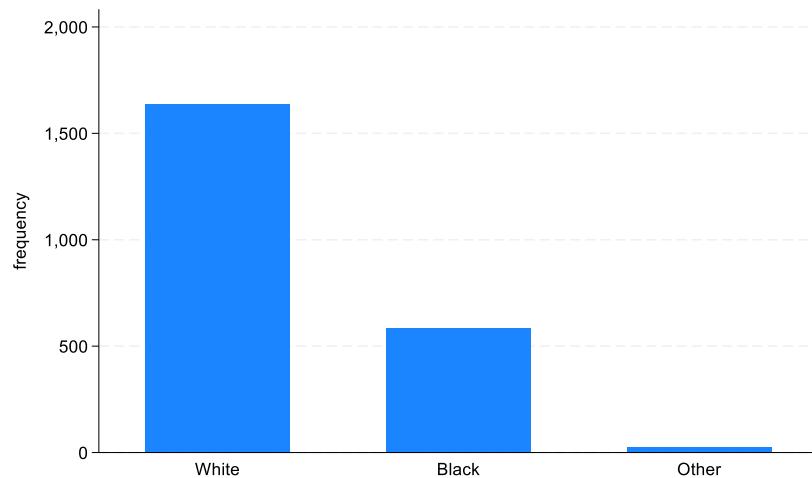
```
sysuse nlsw88.dta, clear
```

## Graphic > Bar Chart

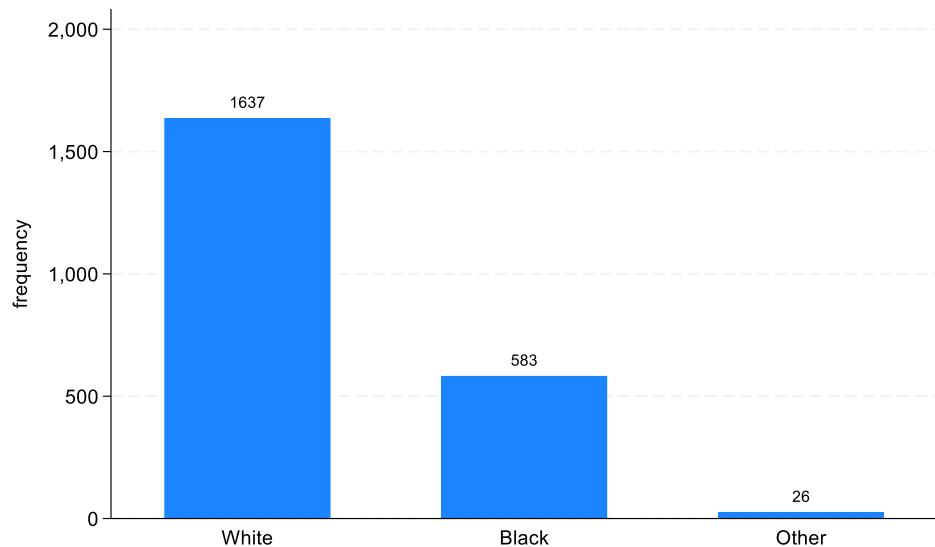
```
graph bar (count)
```



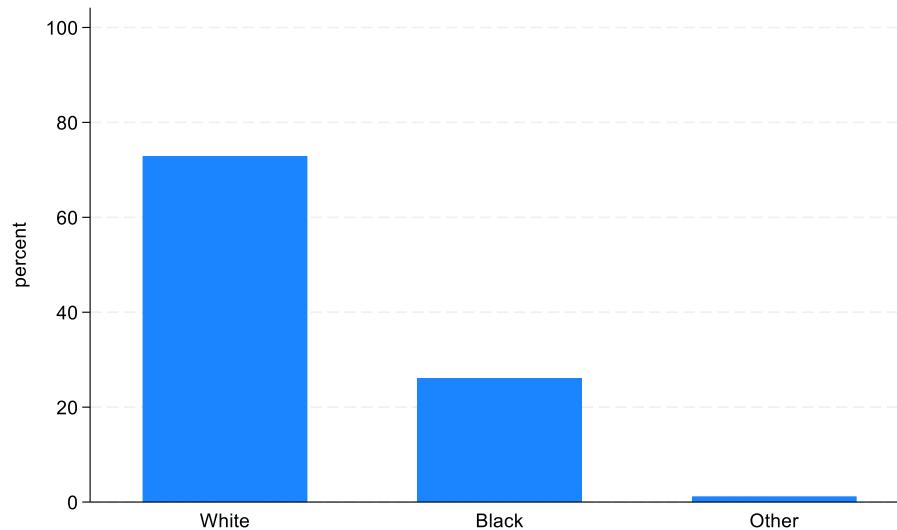
```
graph bar (count), over(race)
```



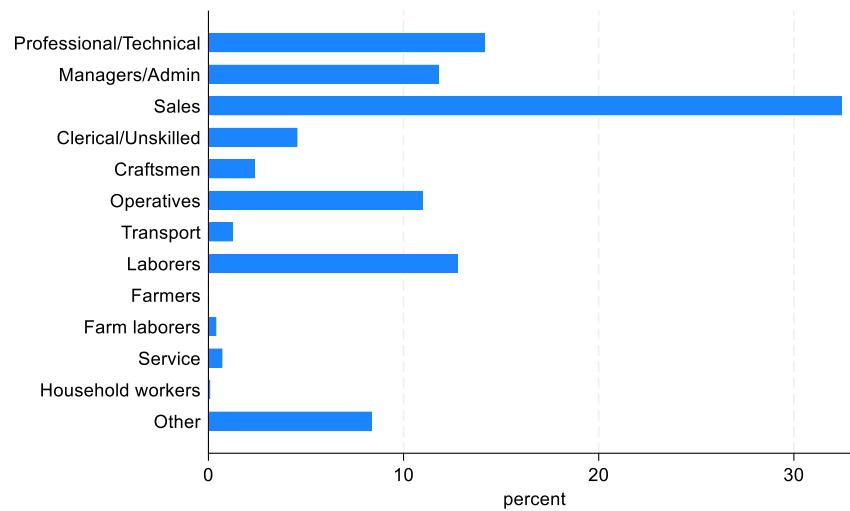
graph bar (count), over(race) blabel(total)



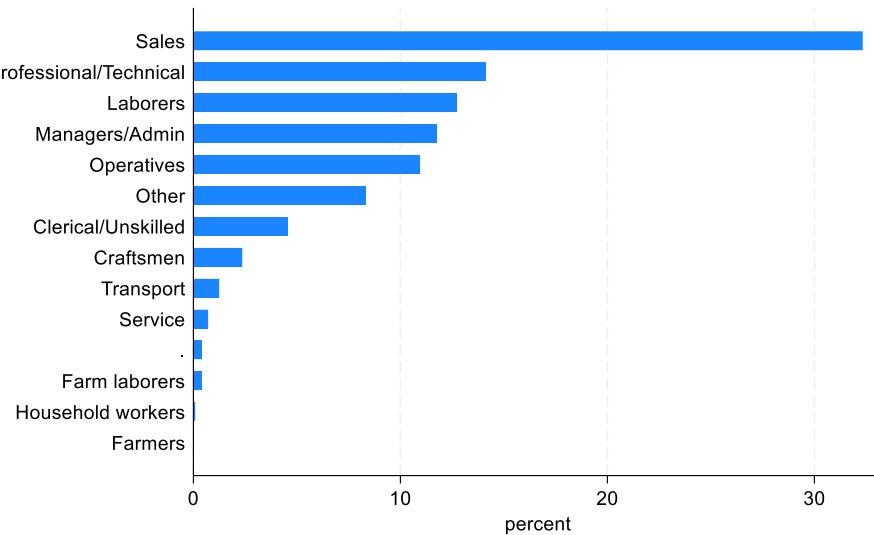
graph bar (percent), over(race) ylabel(0(20)100)



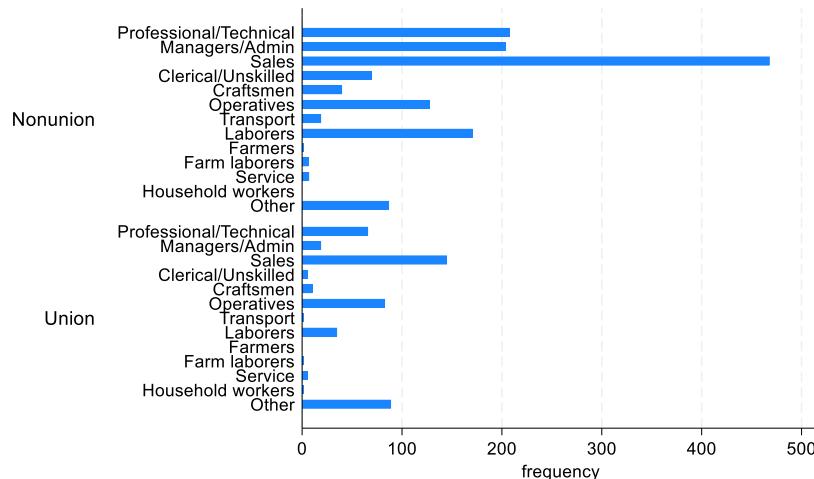
graph hbar (percent), over(occupation)



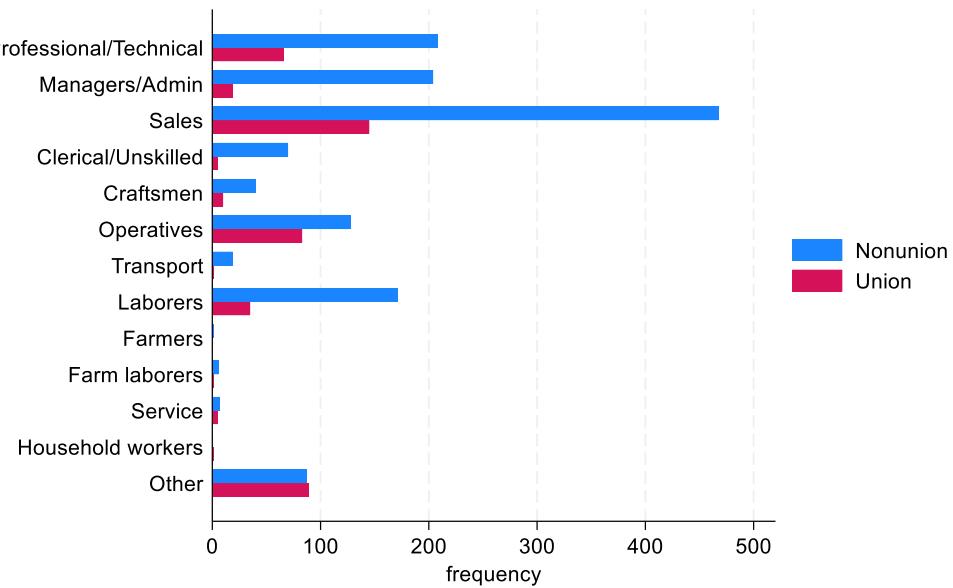
graph hbar (percent), over(occupation, sort(1) descending) missing allcategories



**graph hbar (count), over(occupation) over(union)**



**graph hbar (count), over(union) over(occupation) asyvars**



## แผนภูมิวงกลม

แผนภูมิวงกลม (Pie Chart หรือ Circle Chart) เป็นวงกลมออกเป็นชิ้นส่วนตามสัดส่วนของข้อมูล ยิ่งชิ้นใหญ่ ยิ่งมีค่ามาก

เป็นเครื่องมือแสดงข้อมูลที่ไม่ค่อยมีประสิทธิภาพ แม้ว่าแผนภูมิวงกลมจะ ดึงดูดสายตา, แต่อ่านค่าได้ยาก และเปรียบเทียบข้อมูลไม่ได้ดี

### ปัญหาของแผนภูมิวงกลม

- อ่านค่ายาก → ต้องวัด มุมและพื้นที่ ของชิ้นส่วน
- เหมาะกับข้อมูลที่มีหมวดหมู่ไม่เกิน 5-6 หมวด → ถ้ามีมากกว่านี้ จะอ่านยาก
- ไม่สามารถใช้เปรียบเทียบข้อมูลหลายตัวพร้อมๆ กัน → ต่างจากแผนภูมิแท่งที่สามารถเพิ่มหมวดหมู่ได้
- สามารถบิดเบือนง่าย → ใช้เทคนิค 3D หรือ Exploded Slice ทำให้เข้าใจได้

## ทางเลือกที่ดีกว่าแผนภูมิวงกลม

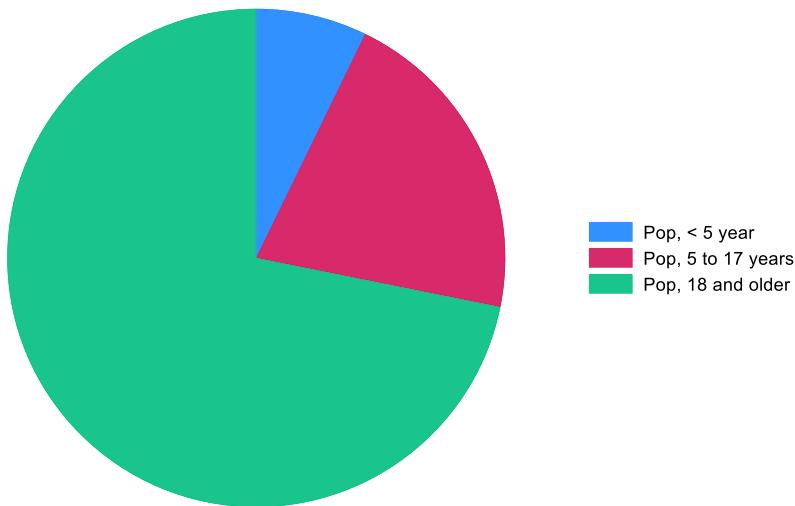
- ใช้แผนภูมิแท่ง (Bar Chart) แทน → อ่านง่ายกว่า, เปรียบเทียบข้อมูลได้ชัดเจนกว่า
- ใช้แผนภูมิแท่งซ้อน (Stacked Bar Chart) → หากต้องการเปรียบเทียบสัดส่วนหลายหมวดหมู่
- ใช้ตาราง (Table) หากต้องการค่าตัวเลขที่แม่นยำ

```
sysuse census.dta, clear
```

Graphics > Pie Chart

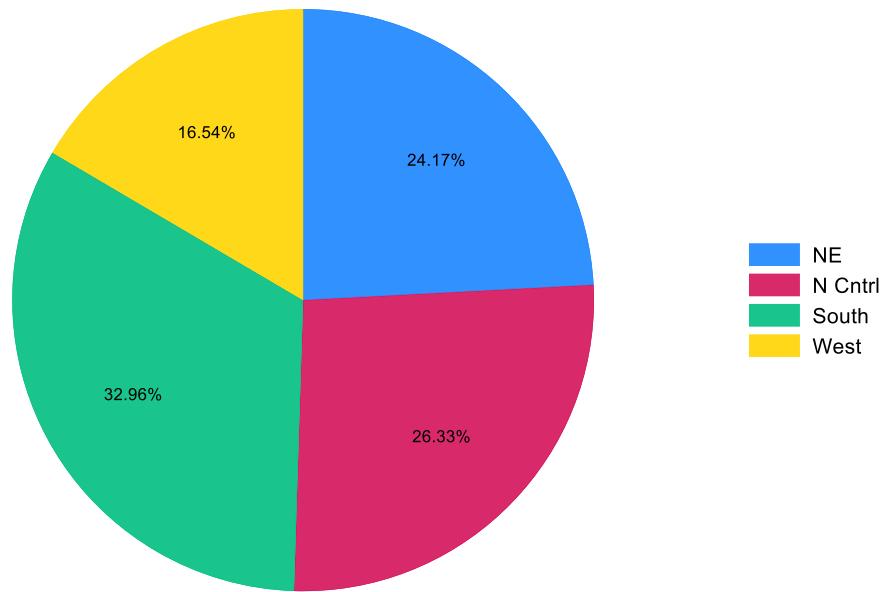
- By variables

```
graph pie poplt5 pop5_17 pop18p
```



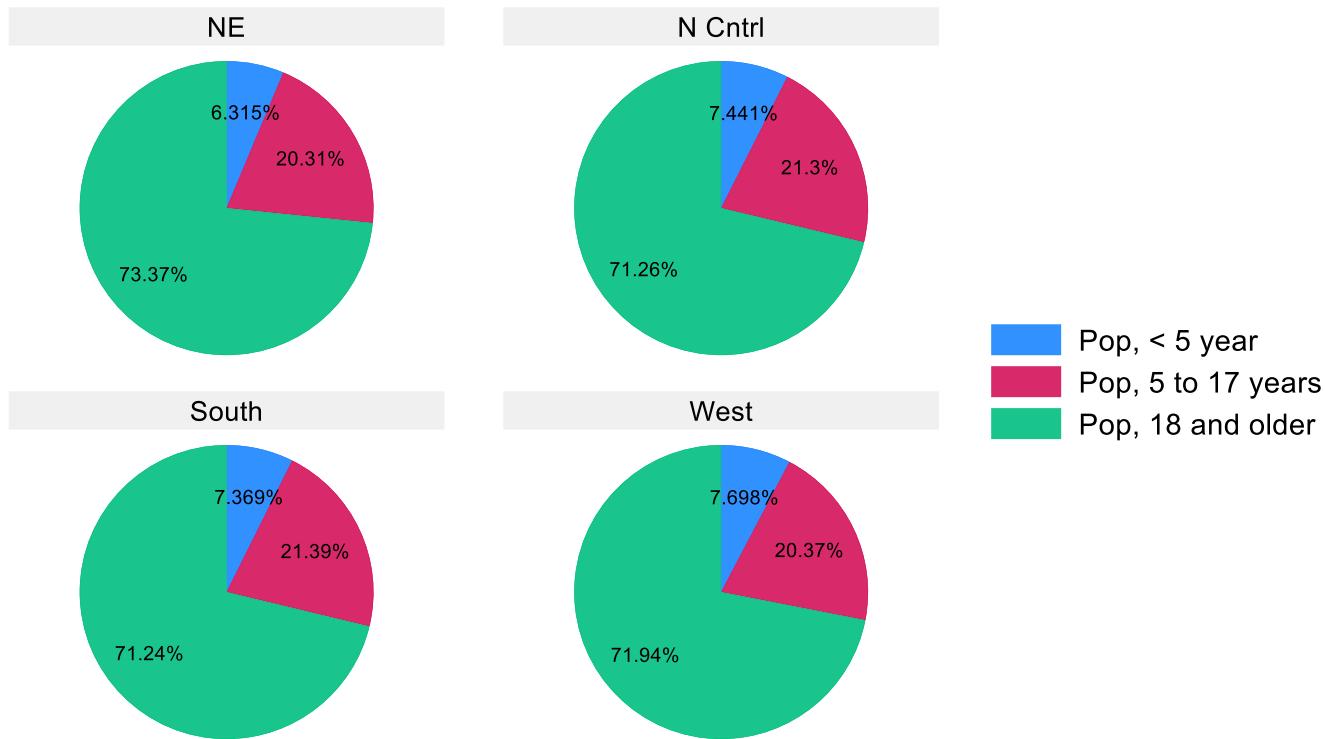
## - By categories

graph pie death, over(region) plabel(\_all percent)



## Multiple pies

```
graph pie poplt5 pop5_17 pop18p, by(region) plabel(_all percent)
```



Graphs by Census region

## **Dot Chart**

Dot Chart (Dot Plot) เป็นวิธีหนึ่งทางเลือกแทนแผนภูมิแท่ง (Bar Chart) โดยแทนที่ แท่ง ด้วย จุด เพื่อแสดงค่าความถี่หรือเปอร์เซ็นต์

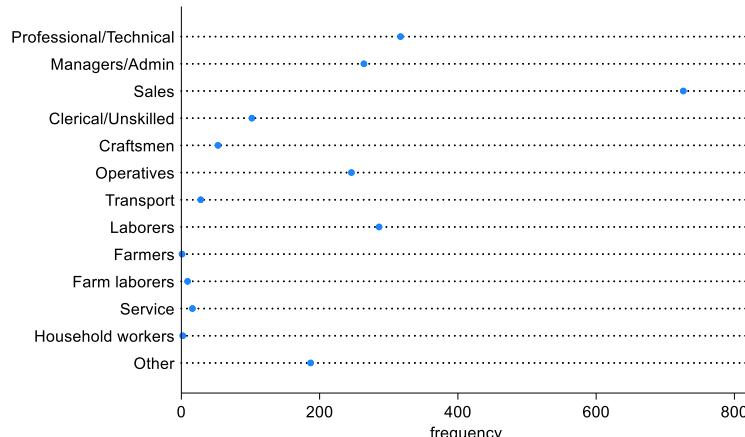
เหมาะสมกับชุดข้อมูลที่มีหมวดหมู่จำนวนมาก ลดความแน่นของกราฟ

ช่วยให้เปรียบเทียบค่าที่ใกล้เคียงกันได้ง่ายขึ้น

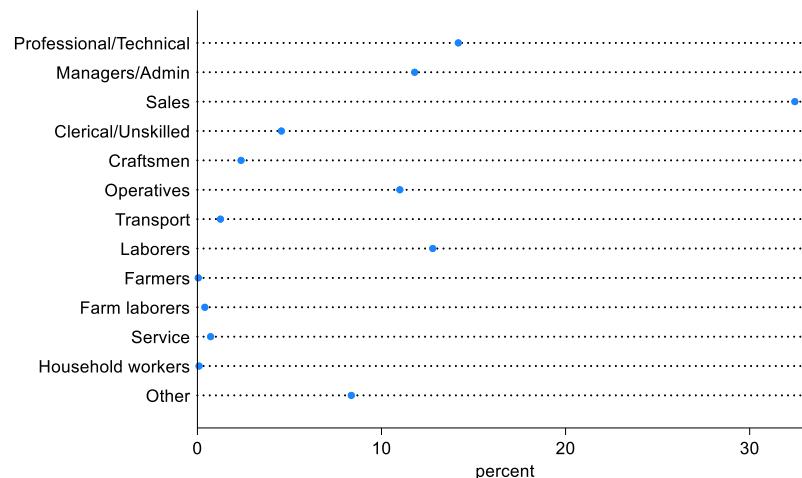
```
sysuse nlsw88.dta, clear
```

## Graphic > Dot Chart

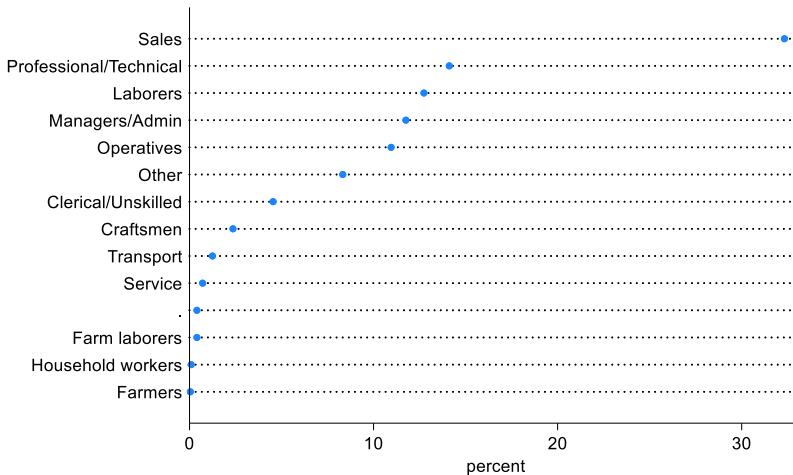
```
graph dot (count), over(occupation)
```



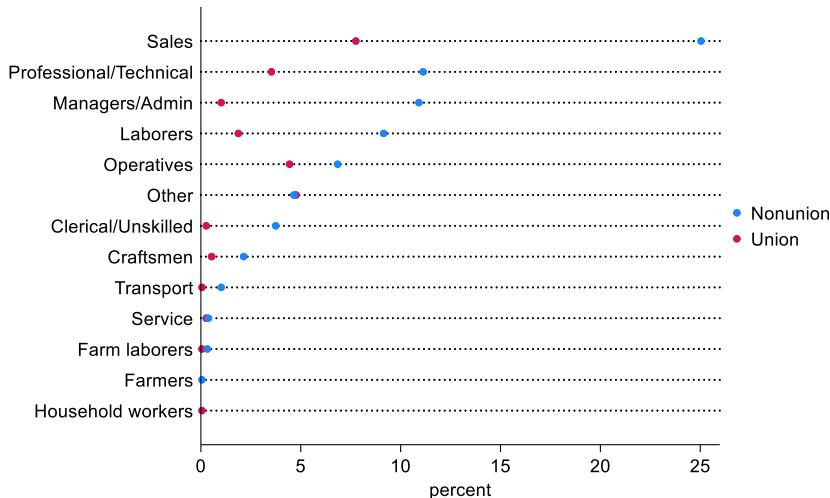
```
graph dot (percent), over(occupation)
```



```
graph dot (percent), over(occupation, sort(1) descending) missing
allcategories
```

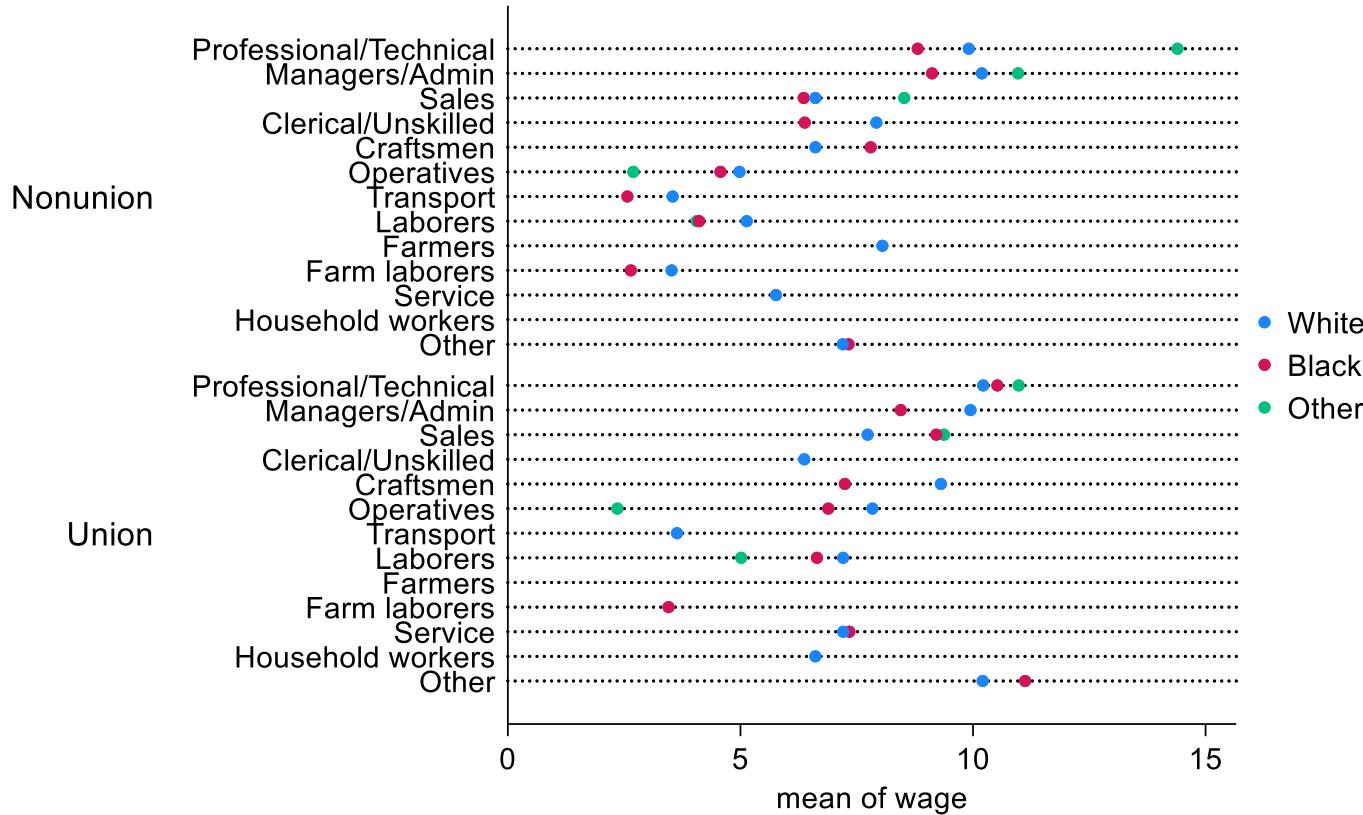


```
graph dot (percent), over(union) over(occupation, sort(1)
descending) asyvars
```

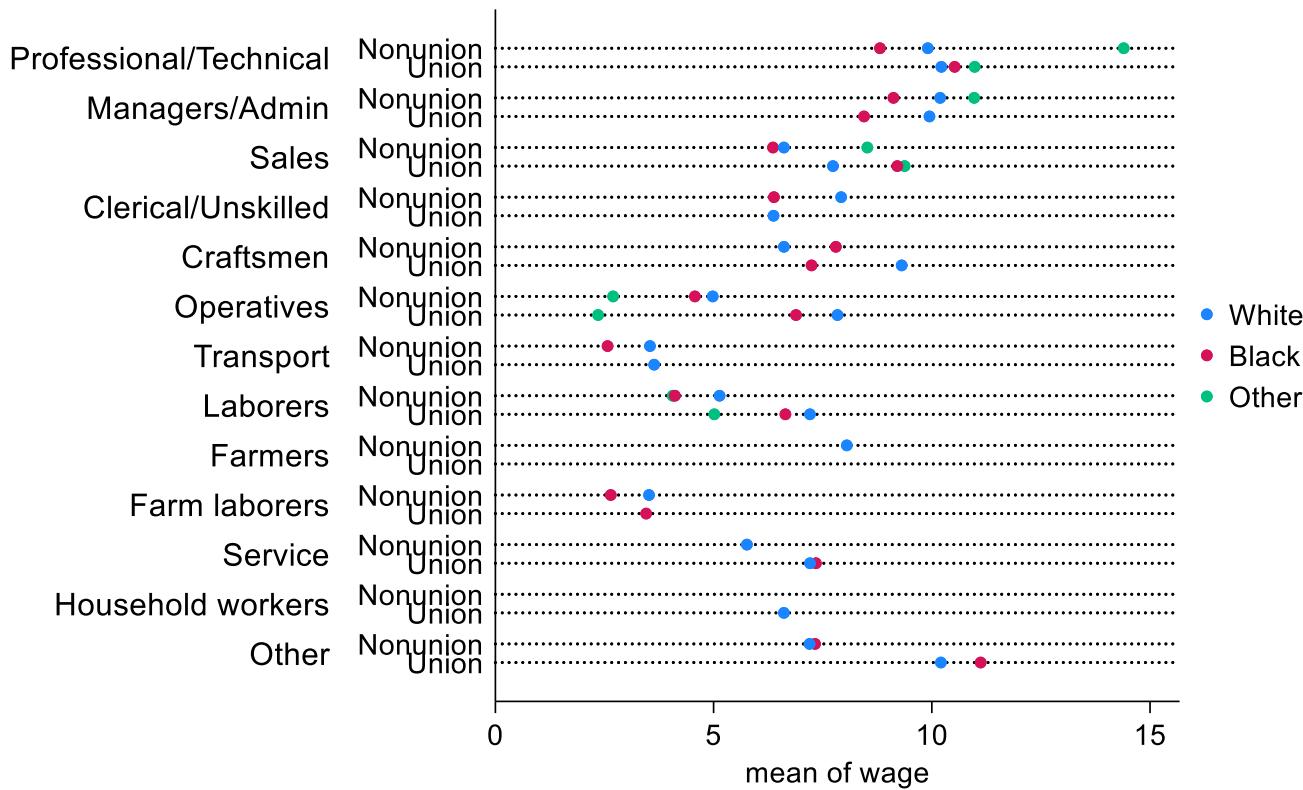


Dot chart over three categories with summary statistics

**graph dot (mean) wage, over(race) over(occupation)  
over(union) asyvars**



```
graph dot (mean) wage, over(race) over(union)
over(occupation) asyvars
```

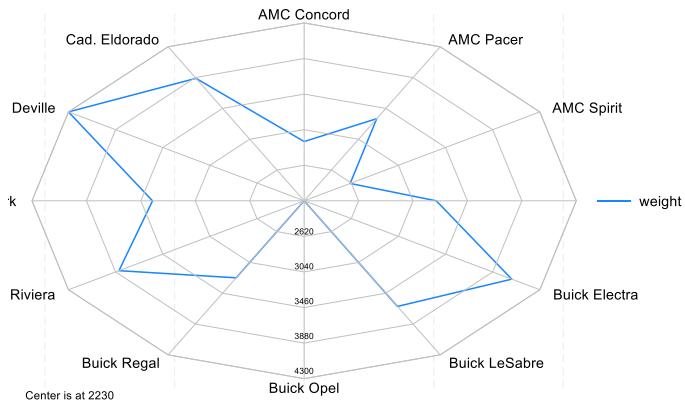


## Radar Plot (Spider Plot)

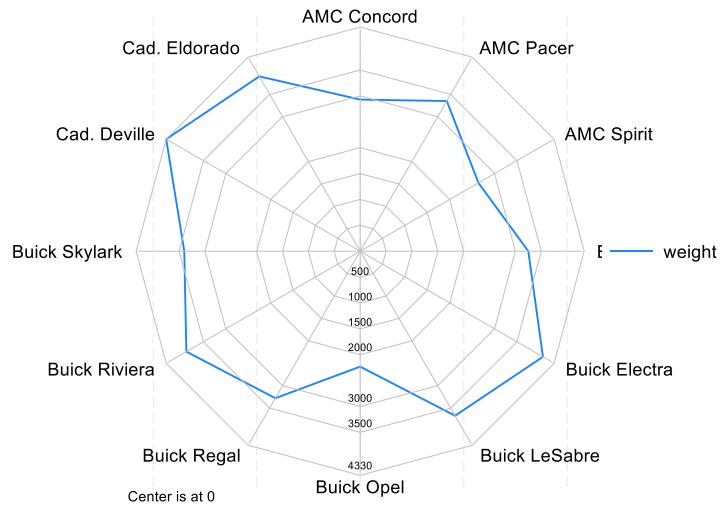
เป็นแผนภูมิวงกลม ที่ใช้เปรียบเทียบหมวดหมู่หลายตัวแปร โดยแต่ละแกน (spoke) แทนหนึ่งหมวดหมู่ และ ความยาวของแกนแสดงค่าสถิติ เช่น เปอร์เซ็นต์ หรือความถี่ จากนั้นเชื่อมต่อจุดข้อมูลทั้งหมดเพื่อสร้างรูปหลายเหลี่ยม

Radar Plot มักใช้ในการวิเคราะห์ข้อมูลหลายมิติ ที่ต้องการเปรียบเทียบคุณลักษณะต่างๆ ระหว่างกลุ่ม

```
sysuse auto.dta, clear  
ssc install radar, replace  
radar make weight in 1/12
```

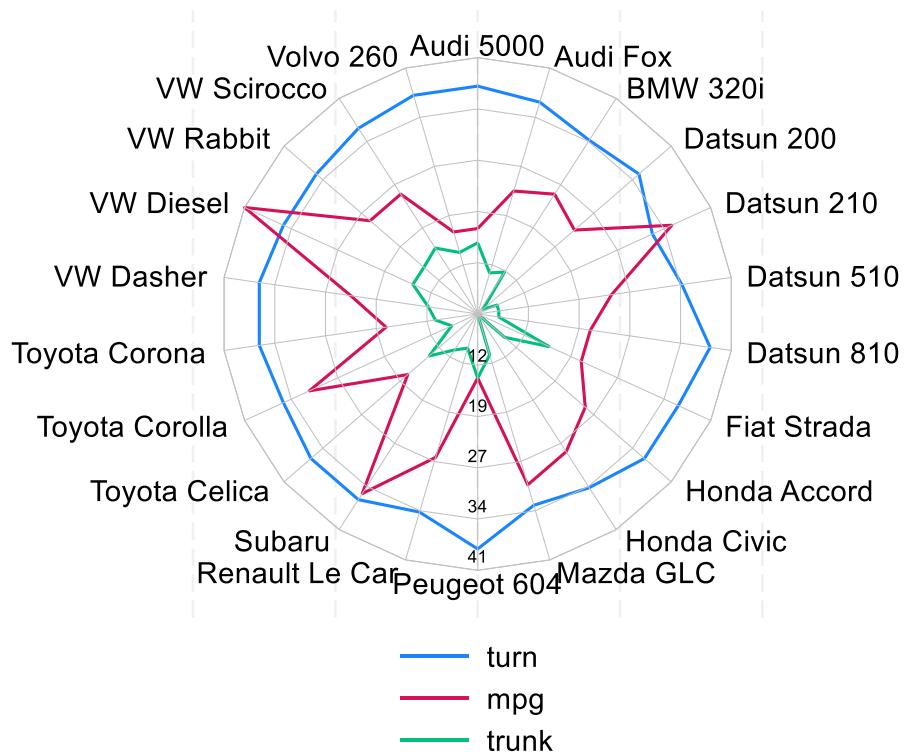


```
radar make weight in 1/12, aspect(1) rlabel(0 500 1000 1500 2000  
3000 3500)
```



Radar plot with multiple variable

**radar make turn mpg trunk if foreign == 1, aspect(1)**



## 5. กราฟสำหรับตัวแปรต่อเนื่องสองตัว (Two Continuous Variables)

---

พิจารณาความสัมพันธ์ระหว่างตัวแปรต่อเนื่องสองตัวแปร

## Scatter Plot

เป็นแผนภูมิที่ใช้แสดงความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ 2 ตัว โดยใช้จุด (dot) ในการplotข้อมูล  
ตัวแปรหนึ่งอยู่บนแกน X และอีกตัวแปรอยู่บนแกน Y

แผนภูมิชนิดนี้ไม่สรุปข้อมูล แต่ให้เห็นแนวโน้มของข้อมูลในรูปแบบดิบ

Scatter Plot หมายถึงอะไร?

- ใช้สำรวจความสัมพันธ์ระหว่างตัวแปรต่อเนื่อง (เช่น รายได้ vs. การบริโภค)
- ใช้ดูแนวโน้ม ความสัมพันธ์ และการกระจายตัวของข้อมูล
- ใช้ค้นหาจุดผิดปกติ (Outliers) และกลุ่มข้อมูล (Clusters)
- สามารถใช้สี, รูปแบบ, หรือป้ายชื่อช่วยแยกกลุ่มข้อมูลได้

## ข้อดี

- แสดงข้อมูลดิบโดยไม่มีการปรับแต่ง → ช่วยให้เห็นรูปแบบจริงของข้อมูล
- ยืดหยุ่นและสามารถขยายได้ → สามารถใช้เป็น Bubble Chart, Heatmap หรือ 3D Scatter Plot
- ช่วยให้เปรียบเทียบกลุ่มข้อมูลได้ง่าย → โดยใช้ สี หรือ รูปแบบจุดต่างๆ
- สามารถเพิ่มเส้นแนวโน้ม (Trend Line) → เพื่อช่วยวิเคราะห์รูปแบบของข้อมูล

## ข้อเสีย

- ถ้าข้อมูลมีจุดเยอะเกินไป อาจอ่านยาก → ข้อมูลซ้อนทับกันจนดูไม่ออก
- ไม่ได้สรุปความสัมพันธ์โดยอัตโนมัติ → ต้องใช้ ค่า Correlation หรือ Regression เพิ่มเติม
- อาจทำให้เข้าใจผิดว่าเป็นความสัมพันธ์เชิงเหตุผล (Causation) → จริงๆ แล้ว เป็นเพียงความสัมพันธ์ (Correlation) เท่านั้น
- ไม่เหมาะสมกับข้อมูลที่เป็นตัวแปรไม่ต่อเนื่อง (Discrete Data) ทั้งสองตัว → ควรใช้ Bar Chart หรือ Dot Chart แทน

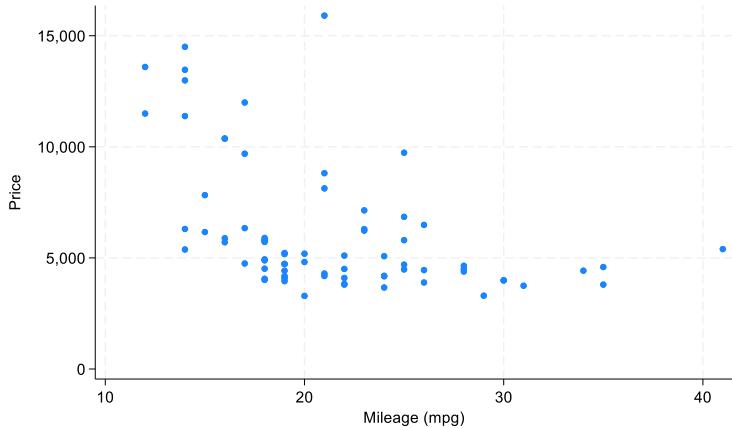
## วิธีแก้ปัญหาข้อมูลแน่นเกินไปใน Scatter Plot:

- ใช้ความโปร่งใส (Transparency/Opacity) → เพื่อให้เห็นจุดที่ซ้อนทับกัน
- ลดขนาดของจุด (Marker Size) → เพื่อลดการทับซ้อนของข้อมูล
- ใช้ Hexbin Plot หรือ Density Plot → แทน Scatter Plot หากมี ข้อมูลมากกว่า 100,000 จุด

```
sysuse auto, clear
```

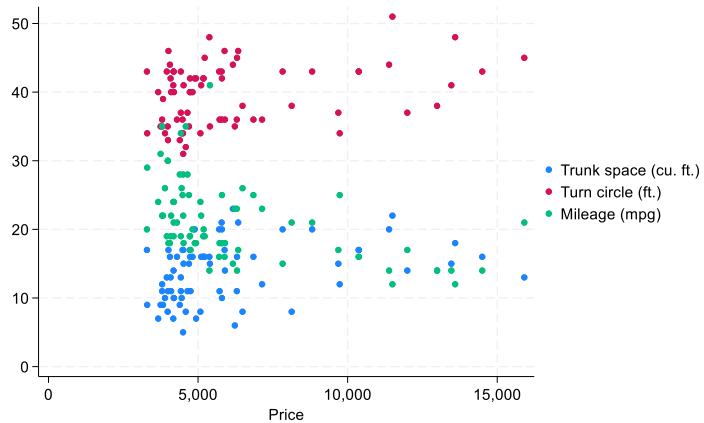
Graphics > Twoway graph (Scatter, line, etc.)

```
twoway (scatter price mpg)
```

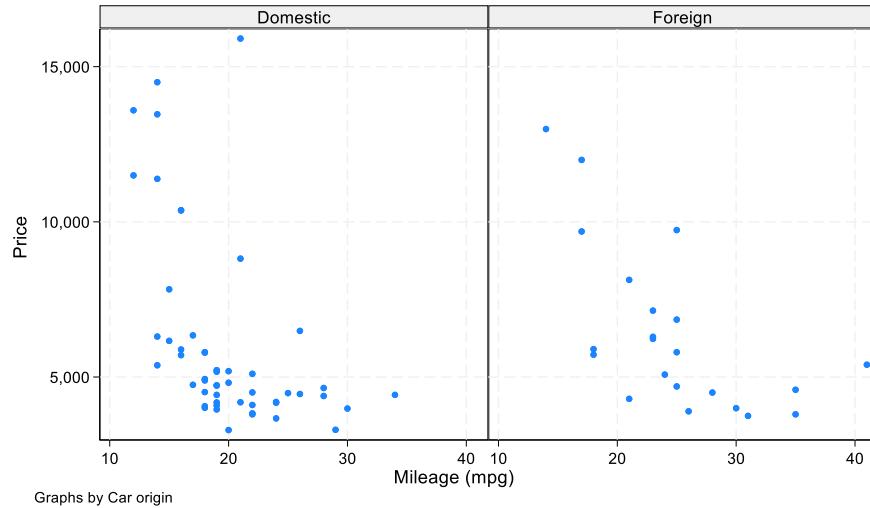


Multiple y-variable against one x-variable

```
twoway (scatter trunk turn mpg price)
```

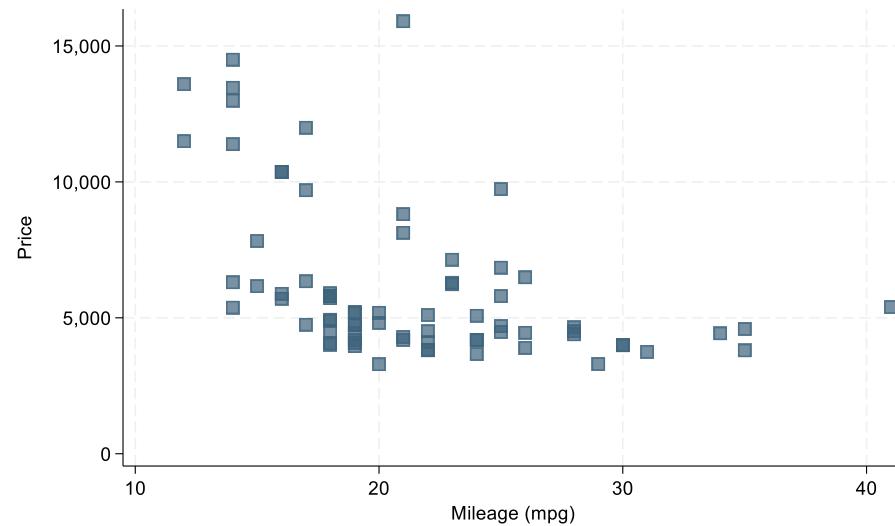


**twoway (scatter price mpg, by(foreign, compact))**

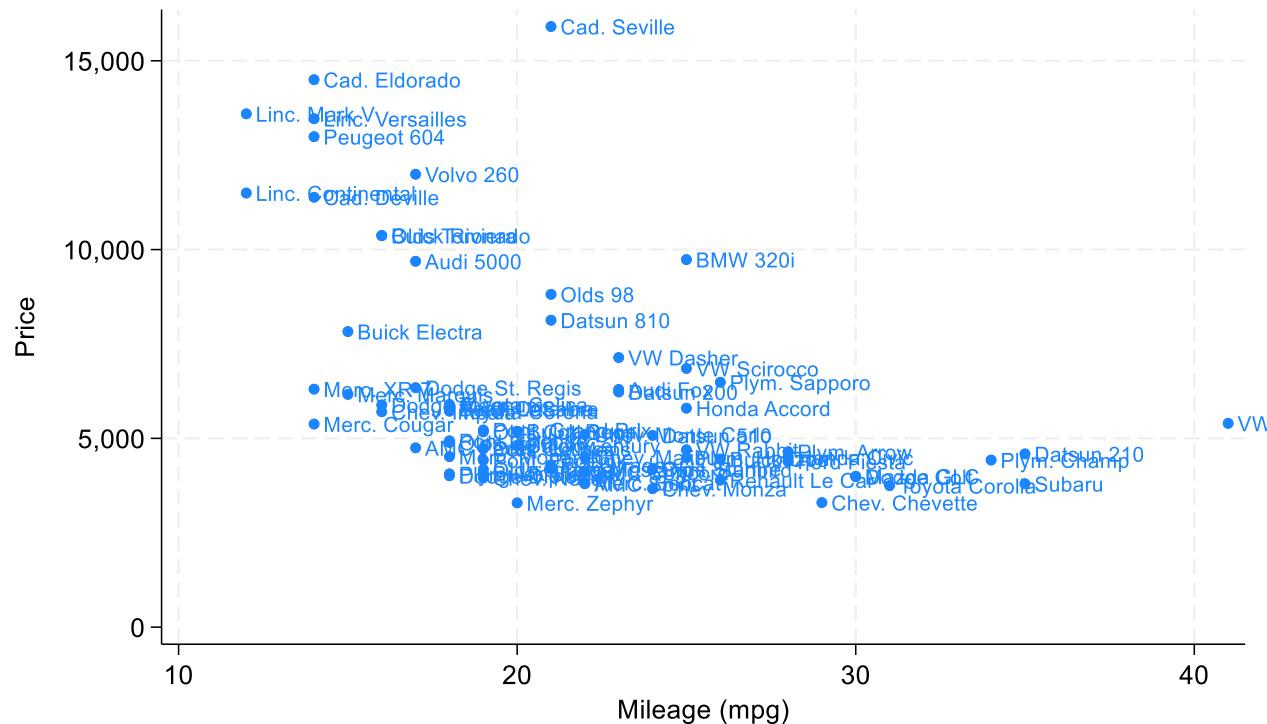


Graphs by Car origin

**twoway (scatter price mpg, msymbol(S) mcolor(edkblue%70) msiz(\*2))**

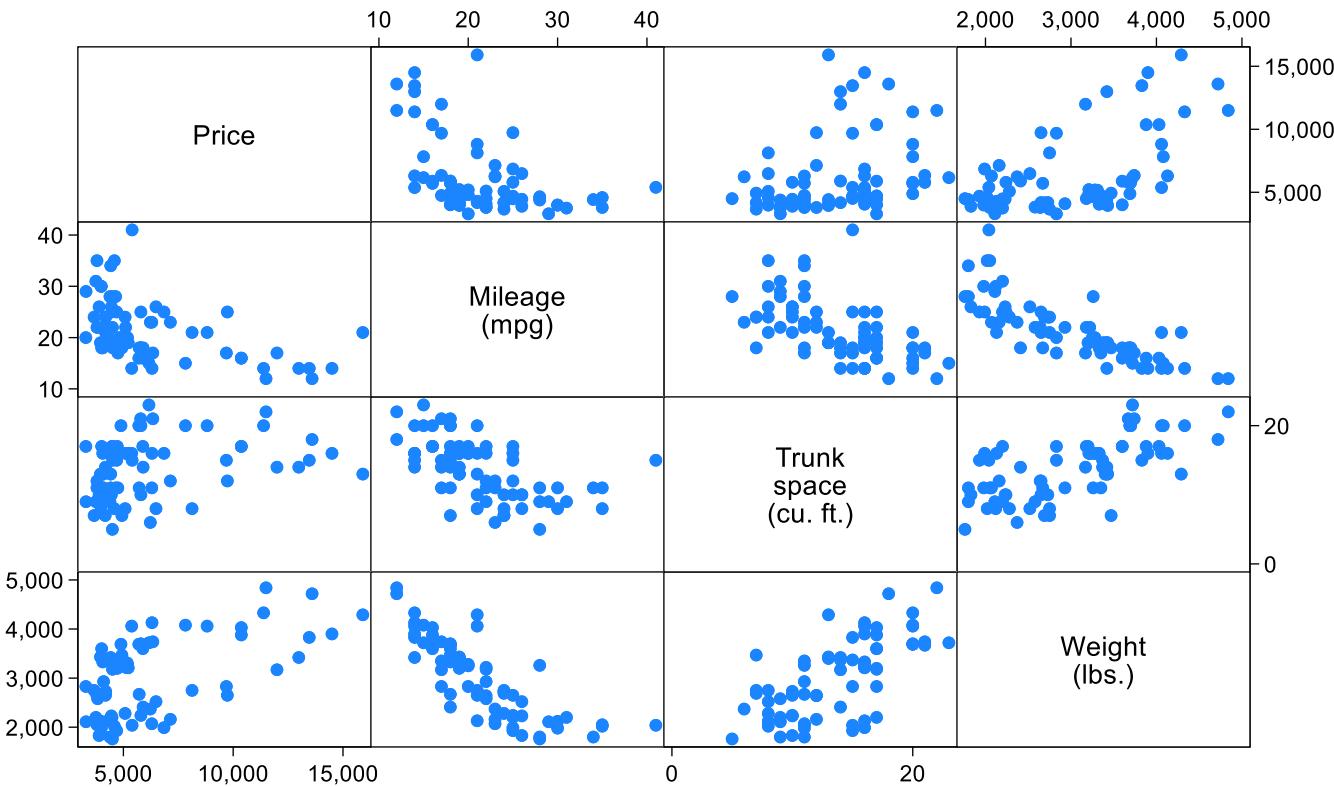


```
twoway (scatter price mpg, mlabel(make))
```



## Scatterplot Matrix

**graph matrix price mpg trunk weight**



## Heat Map

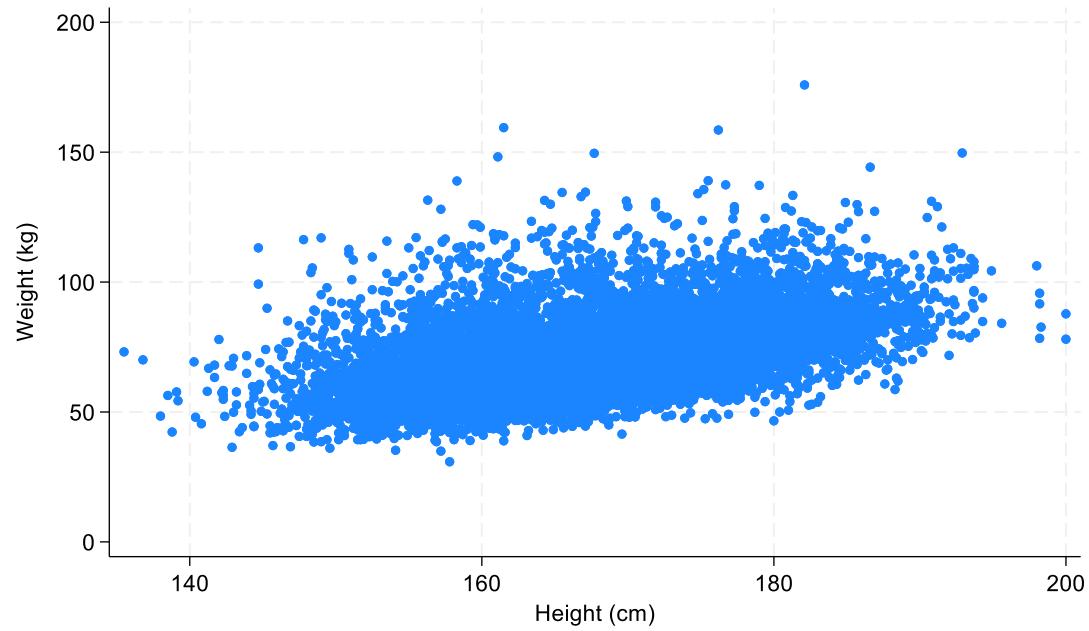
Heat Map หรือ Heat Scatter Plot เป็นวิธีการแสดงผลทางเลือกแทน Scatter Plot โดย亥ม่ากับข้อมูลขนาดใหญ่ที่มีจุดข้อมูลมากจนซ้อนทับกันมากเกินไป

Heat Map จะแบ่งข้อมูลออกเป็นกริด (bins) และใช้ สีแสดงความหนาแน่นของข้อมูลในแต่ละช่อง

วิธีปรับปรุง Heat Map ให้ดูง่ายขึ้น:

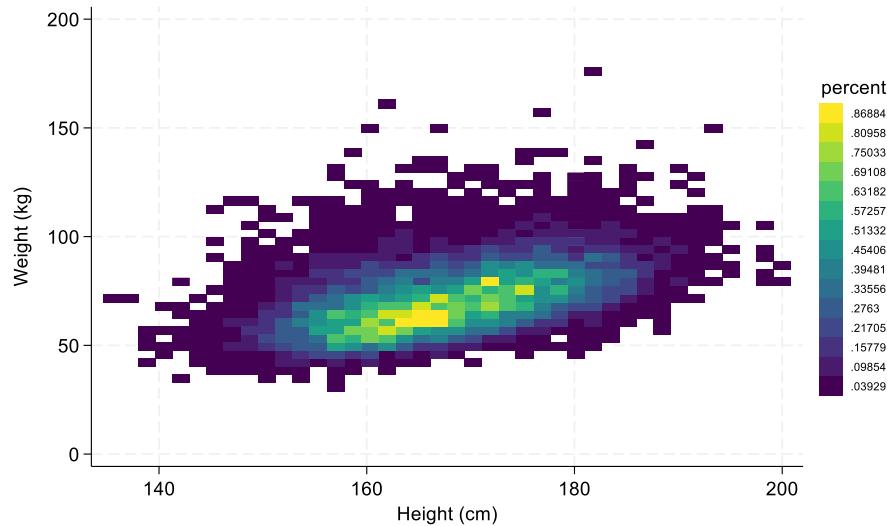
- เลือกสเกลสีที่亥ม่าสม → สีที่แตกต่างกันชัดเจนจะช่วยให้ เห็นความหนาแน่นของข้อมูลได้ดีขึ้น
- ลองปรับขนาด bin ให้亥ม่าสม → Bin ใหญ่เกินไป = ข้อมูลดูหยาบ, Bin เล็กเกินไป = มี Noise มากเกินไป
- เพิ่มเส้นแนวโน้ม (Trend Line) ลงไปด้วย → ช่วยให้เข้าใจความสัมพันธ์ระหว่างตัวแปรได้ดีขึ้น
- ใช้ความโปร่งใส (Transparency/Opacity) → ลดปัญหาสีทึบเกินไป

```
webuse nhanes2.dta, clear  
tw (sc weight height)
```

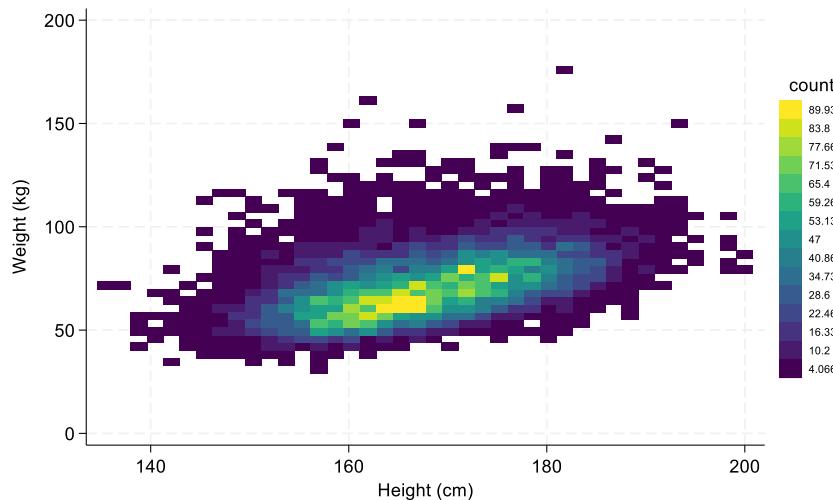


```
ssc install heatplot, replace
```

## heatplot weight height



## heatplot weight height, statistic(count)



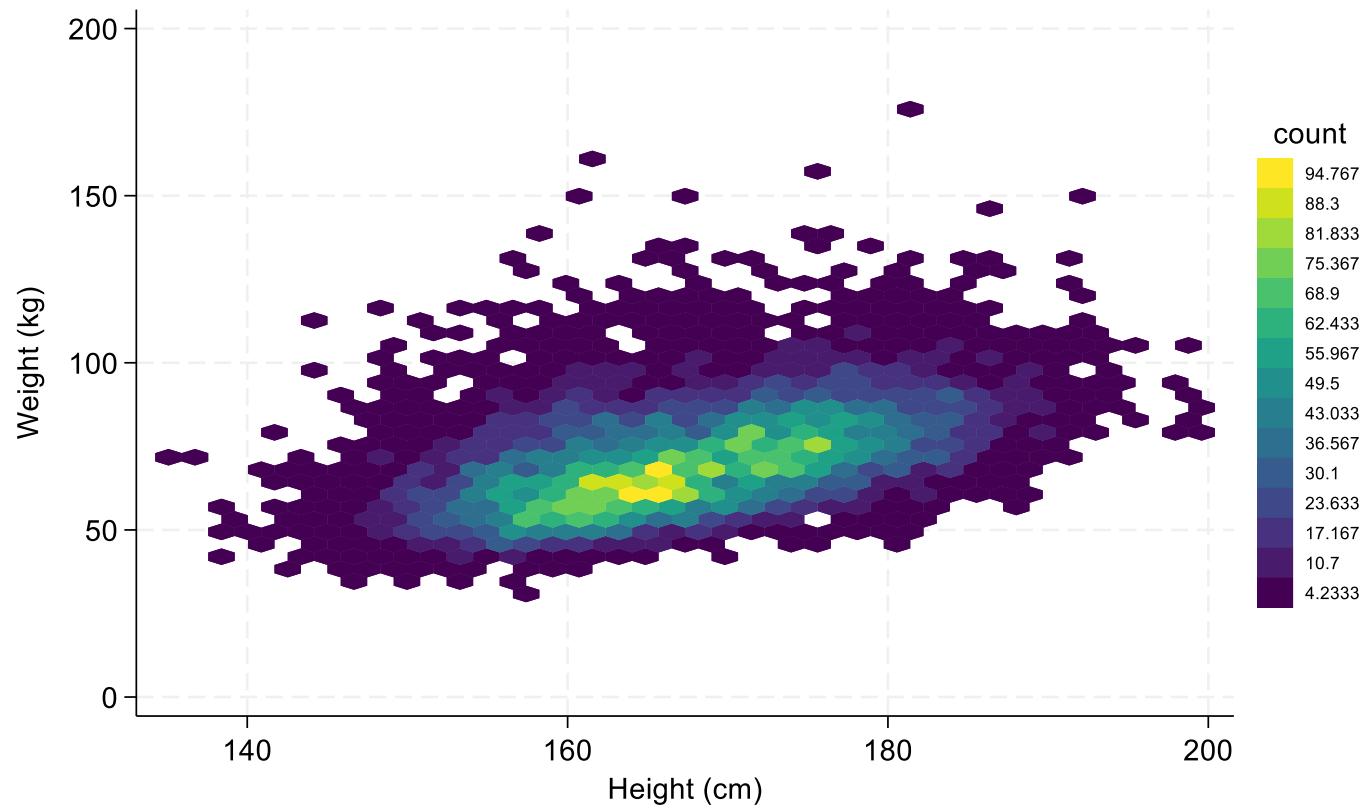
## Hexbin Plot

Hexbin Plot หรือ Hexagonal Heat Map เป็นทางเลือกของ Heat Map แบบตารางสี่เหลี่ยมที่ใช้หกเหลี่ยมแทนสี่เหลี่ยมในการแบ่งพื้นที่ของ Scatter Plot เพื่อให้แสดงแนวโน้มของข้อมูลได้ดียิ่งขึ้น โดยเฉพาะข้อมูลที่มีการกระจายตัวเป็นแนวทแยง (Diagonal Relationships)

### ทำไมต้องใช้ Hexbin แทน Square Heat Map?

- สี่เหลี่ยมมีเพียง 4 ทิศทาง → ข้อมูลเชื่อมโยงกันแค่ แนวตั้งและแนวอน
- หกเหลี่ยมมี 6 ทิศทาง → จับแนวโน้มในแนวทแยงมุมได้ดีกว่า
- หกเหลี่ยมคล้ายวงกลมมากกว่า → กลุ่มข้อมูลมีการรวมตัวที่ เป็นธรรมชาติมากขึ้น
- สร้างภาพที่นุ่มนวลกว่าสี่เหลี่ยม → ลดปัญหาภาพดูเป็นตารางแข็งทือ

**heatplot weight height, statistic(count) hex**



## เส้นแนวโน้ม

เป็นเครื่องมือที่ช่วยให้เราเข้าใจความสัมพันธ์ระหว่างตัวแปรต่อเนื่อง 2 ตัว ใน scatter plot ได้ง่ายขึ้น

สามารถแบ่งออกเป็น 2 ประเภทหลักคือ

- Parametric (กำหนดสมการชัดเจน)
- Non-Parametric (ปล่อยให้ข้อมูลกำหนดรูปแบบเอง)

แสดง ความสัมพันธ์เชิงบวก (Positive) หรือ เชิงลบ (Negative) ระหว่างตัวแปร X และ Y

ลดความซับซ้อนของข้อมูลดิบ และช่วยให้เข้าใจความสัมพันธ์ได้รวดเร็ว

## สมการatyตัว (Parametric)

- สมการเส้นตรง:  $Y = a + bX$
- สมการกำลังสอง:  $Y = a + bX + cX^2$

## สมการไม่atyตัว (Nonparametric)

- ไม่ใช้สมการatyตัว แต่คำนวณค่าเฉลี่ยของข้อมูลในกลุ่มเดียวกัน
- ใช้ Bandwidth ในการคำนวณค่าเฉลี่ยของช่วงข้อมูล

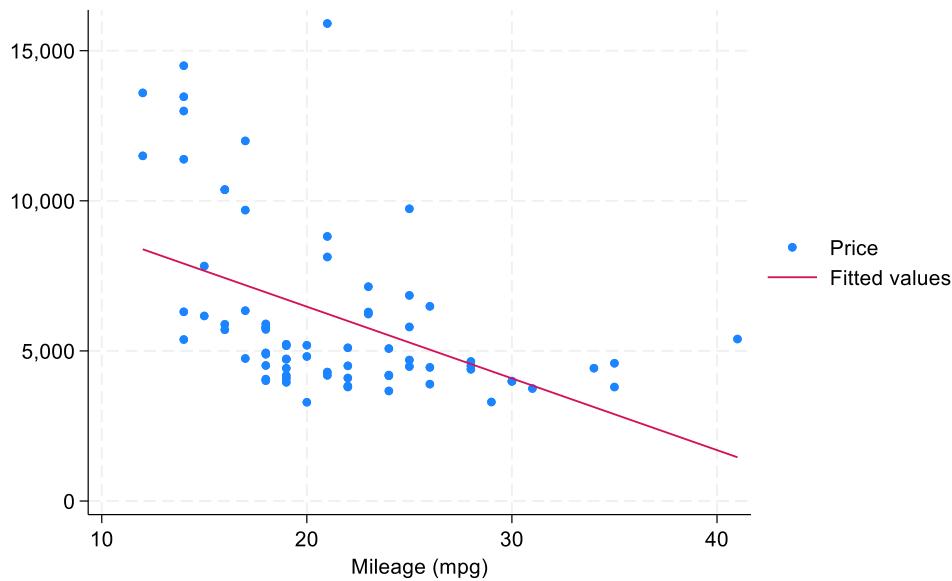
## เมื่อใดที่ควรใช้เส้นแนวโน้ม

- ข้อมูลมีความสัมพันธ์ที่ชัดเจน
- ต้องการสรุปแนวโน้มของข้อมูล

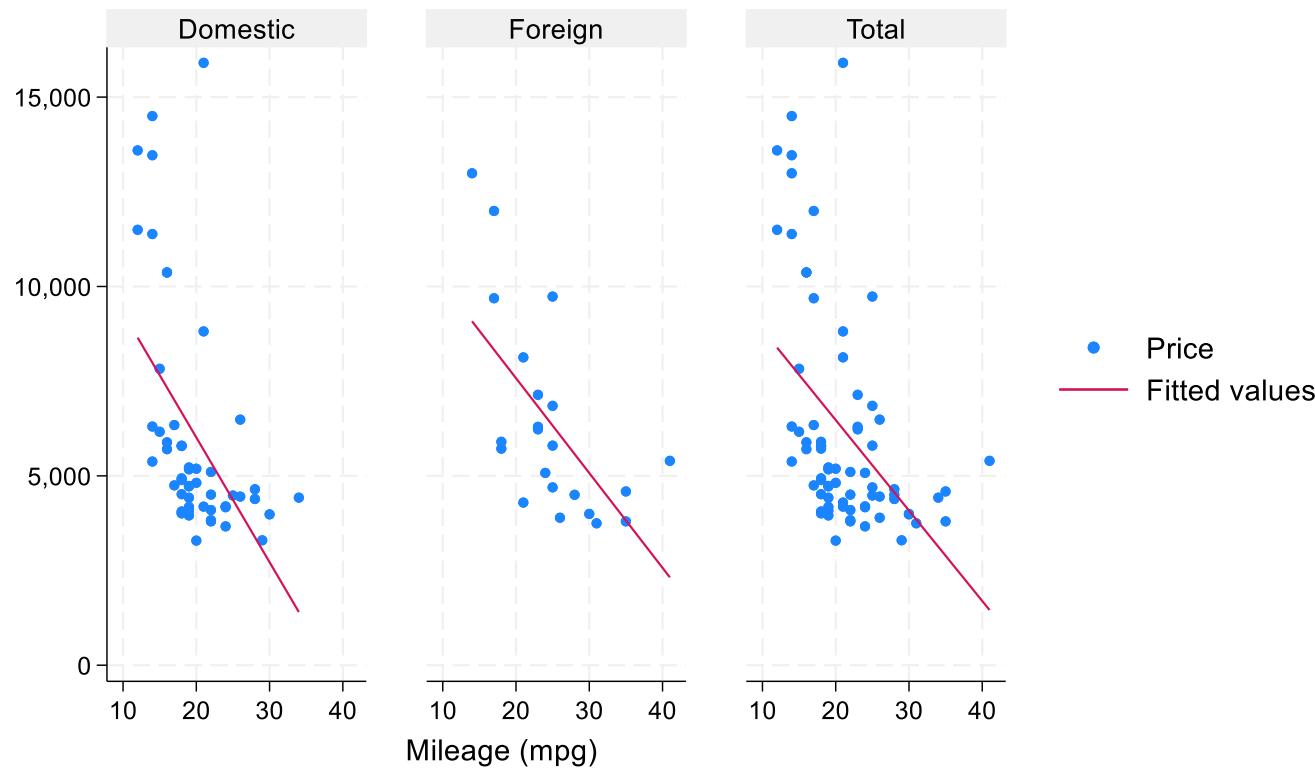
## หลีกเลี่ยงเมื่อ

- ข้อมูลเป็นตัวแปรไม่ต่อเนื่อง (discrete)
- ไม่มีแนวโน้มที่ชัดเจน

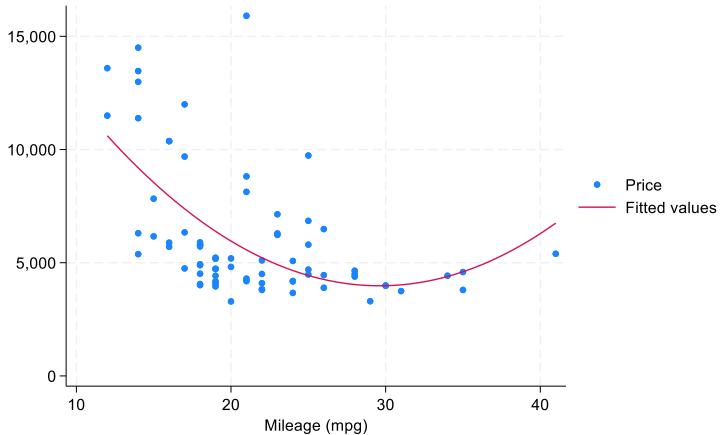
**two way (scatter price mpg) (lfit price mpg)**



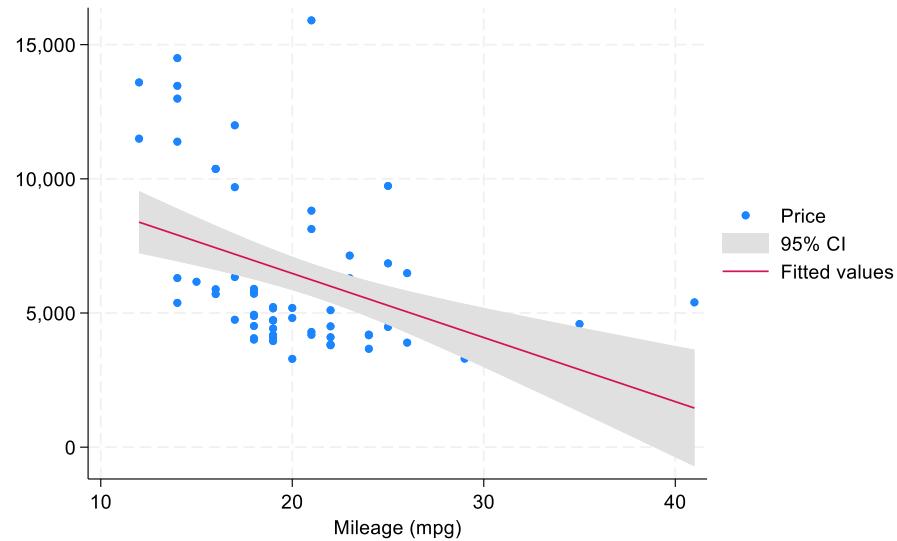
```
tw (scatter price mpg) (lfit price mpg, by(foreign, total row(1)))
```



**tw (scatter price mpg) (qfit price mpg)**



**tw (scatter price mpg) (lfitci price mpg)**



## กราฟเส้น

Line plot เป็นกราฟที่ใช้กันมากในการวิเคราะห์ข้อมูล อนุกรมเวลาและแนวโน้มต่างๆ โดยเส้นที่เชื่อมต่อ กันช่วยให้เห็นรูปแบบของข้อมูลได้ชัดเจนกว่ากราฟแบบจุด (scatter plot)

เหมาะสมสำหรับข้อมูลที่มี ลำดับชัดเจน เช่น วัน, เดือน, ปี

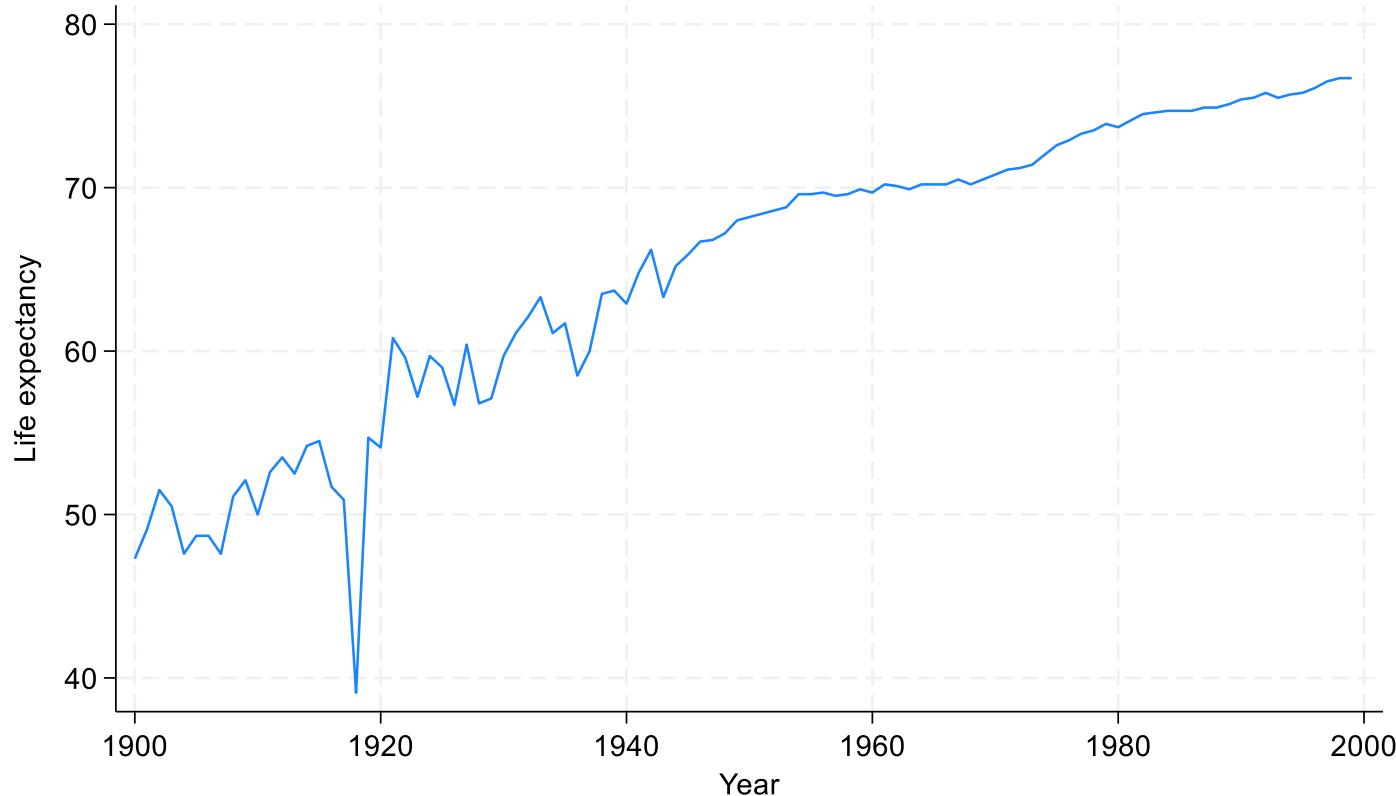
เปรียบเทียบห蓝图ณ์ได้ง่าย (ใช้สีที่แตกต่างกัน กรณีกลุ่มน้อย และเส้นประ กรณีห蓝图ณ์)

ไม่ควรใช้เมื่อ:

- ข้อมูลไม่มีลำดับที่ชัดเจน → กราฟจะดูยุ่งเหยิง
- ข้อมูลมีการเปลี่ยนแปลงขึ้นลงมาก → อาจดู แหลมและอ่านยาก

```
sysuse uslifeexp, clear
```

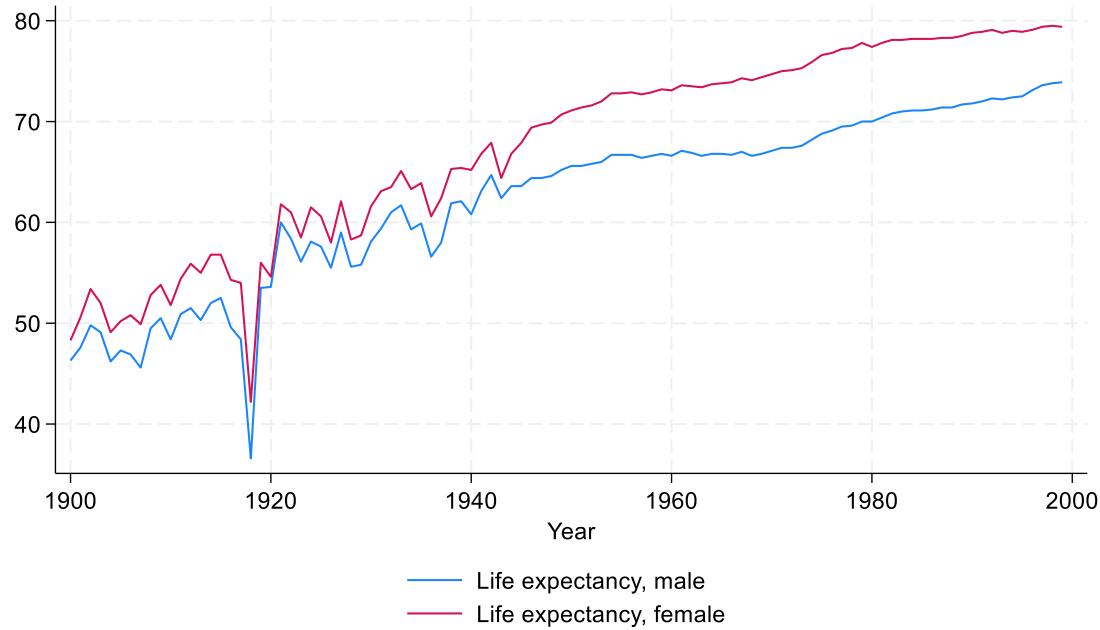
```
tw (line le year)
```



แจ้งให้ Stata ทราบว่าเป็นข้อมูลอนุกรมเวลา

**tsset year**

**twoway (tsline le\_male) (tsline le\_female)**



## แผนภาพพื้นที่

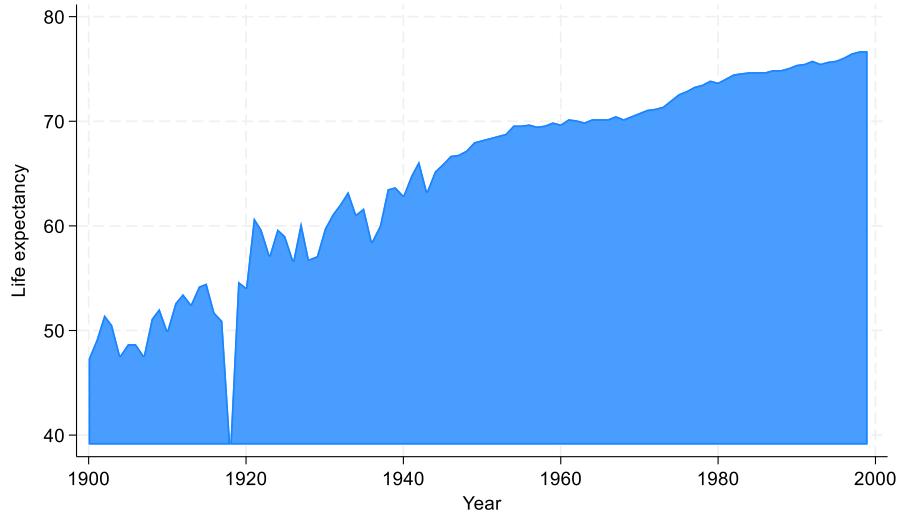
Area Plot เป็นรูปแบบหนึ่งของแผนภูมิเส้น (Line Plots) ที่ช่วยเพิ่มความชัดเจนของแนวโน้มข้อมูล โดยการเน้นพื้นที่ใต้กราฟ

แผนภาพประเภทนี้มีประโยชน์อย่างมากเมื่อใช้กับข้อมูลอนุกรมเวลา (Time Series Data) หรือข้อมูลที่มีความผันผวนสูง

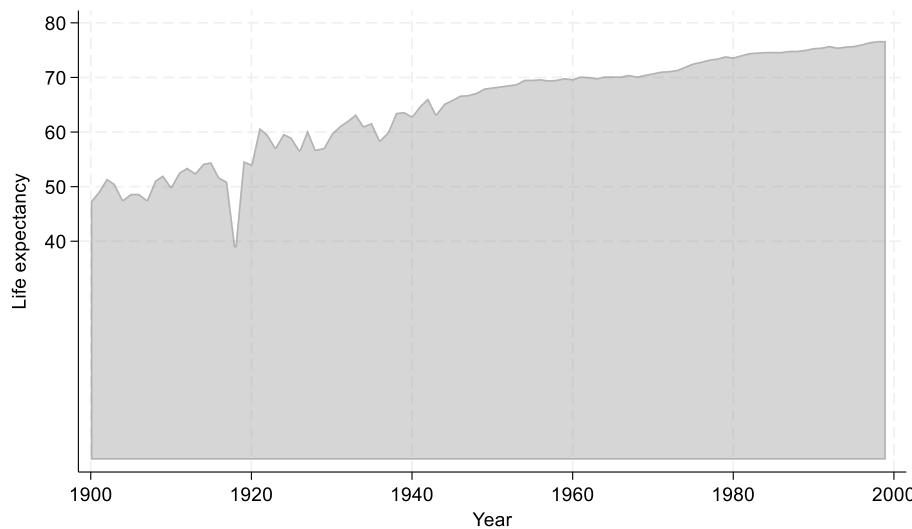
## ทางเลือกอื่น ๆ

- แผนภาพแท่ง (Bar Plot / Column Plot): ใช้แท่งแนวตั้งแทนการเติมสีใต้เส้นกราฟ
- แผนภาพเส้นแหลม (Spike Plot): คล้ายแผนภาพเส้น แต่ใช้เส้นแหลมบาง ๆ แทนแท่งหนา
- แผนภาพเส้นเชื่อม (Dropline Plot): คล้ายแผนภาพเส้นแหลม แต่เพิ่มจุดข้อมูล (Data Markers) เพื่อให้เห็นค่า

two way (area le year)



tw (area le year, base(0) color(gs8%40))



## 6. กราฟสำหรับตัวแปรไม่ต่อเนื่องสองตัว (Two Discrete Variables)

---

แทนที่จะใช้ตารางสองทาง (Two-way Table)

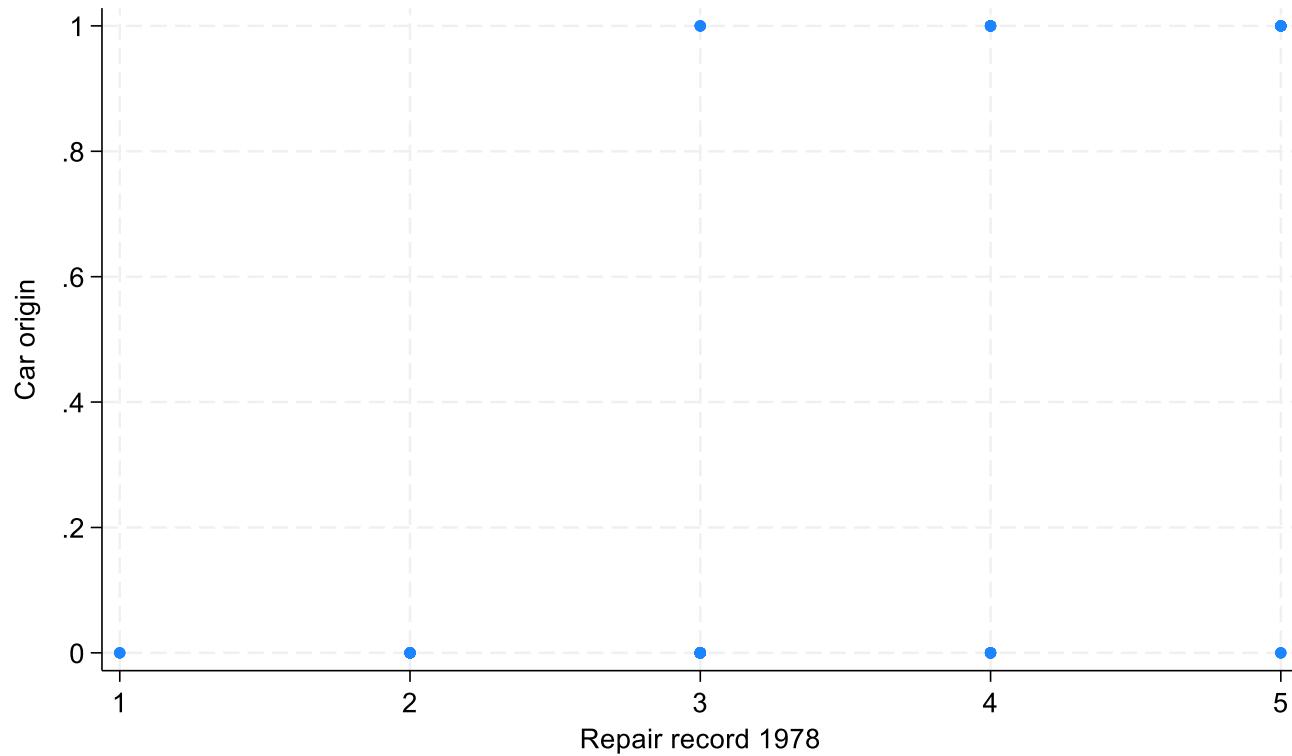
## Jitter Plot

Jittered Scatter Plot เป็น แผนภูมิแบบกระจาย (Scatter Plot) ที่ถูกปรับแต่ง เพื่อช่วยแสดงข้อมูลตัวแปรที่เป็นประเภทไม่ต่อเนื่อง (Discrete Data) โดยการสุ่มเลื่อนจุดข้อมูลออกจากค่าจริงเล็กน้อย เพื่อป้องกันปัญหาจุดทับซ้อนกันมากเกินไป

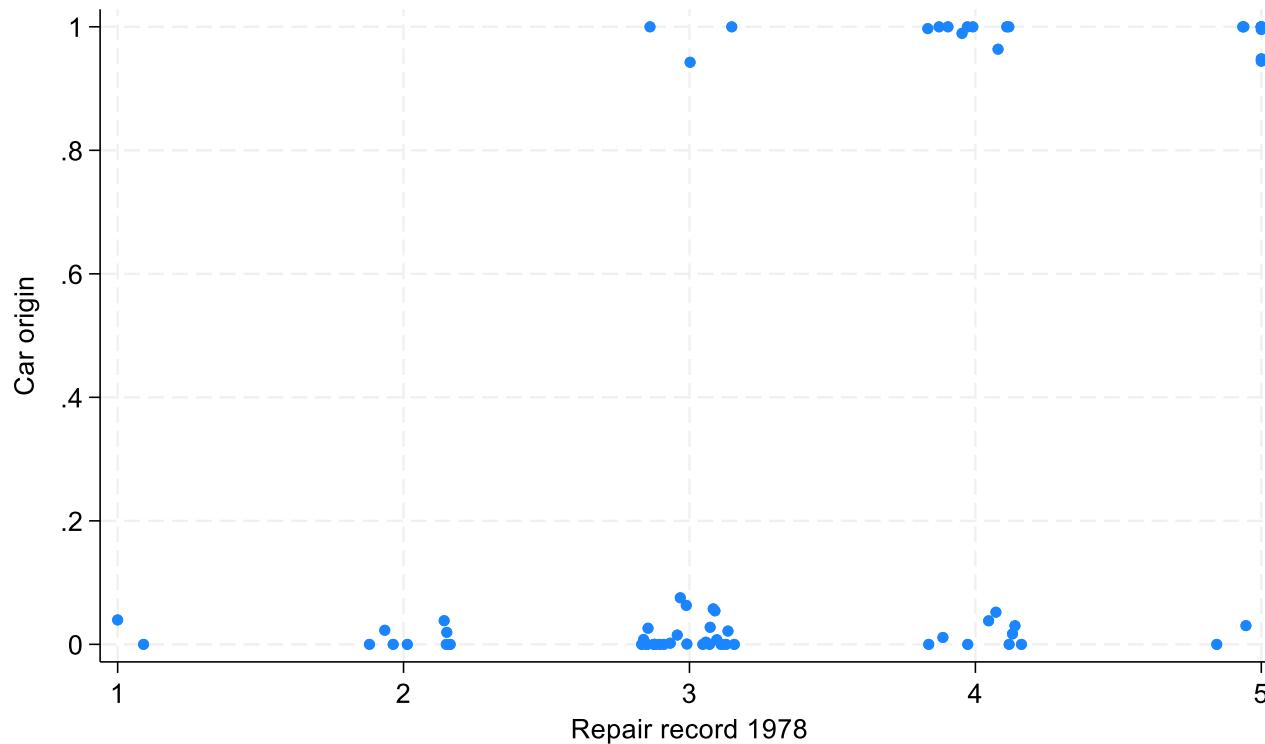
Jitter Plot หมายความว่าอะไร?

- เมื่อทั้ง X และ Y เป็นตัวแปรที่ไม่ต่อเนื่อง (Discrete Variables) และ Scatter Plot ทั่วไปทำให้จุดซ้อนทับกันมากเกินไป
- เมื่อเปรียบเทียบตัวแปร 2 ตัวที่เป็นหมวดหมู่ เช่น กลุ่มอุตสาหกรรม vs อาชีพ
- เมื่อมีข้อมูลแบบกึ่งต่อเนื่อง
- เมื่อชุดข้อมูลมีขนาดเล็กหรือปานกลาง

```
sysuse auto, clear  
tw (scatter foreign rep78)
```



```
tw (scatter foreign rep78, jitter(12))
```



## แผนภูมิตาราง (Table Plot)

แผนภูมิตาราง (Table Plot) เป็น รูปแบบหนึ่งของการแสดงข้อมูลในตารางแบบภาพ โดยใช้ แถบแท่ง (Bar Chart) ที่ถูกจัดรูปแบบพิเศษ เพื่อทำให้การอ่านข้อมูลง่ายขึ้น

แผนที่จะใช้ตัวเลขดิบในตารางสองทาง ซึ่งอาจอ่านยากและใช้เวลานานในการเปรียบเทียบ

### ข้อดี

- สื่อสารข้อมูลได้เร็วขึ้น – ไม่ต้องอ่านค่าตัวเลขทั้งหมด
- ช่วยให้เห็นแนวโน้มข้อมูลได้ง่ายขึ้น
- ลดความซับซ้อนของตารางตัวเลข
- สามารถเพิ่มรายละเอียดเพิ่มเติม เช่น การแสดงเปอร์เซ็นต์แทนค่าจริง

### ข้อเสีย

- เสียความแม่นยำของข้อมูลเชิงตัวเลข – ไม่สามารถอ่านค่าที่แน่นอนได้โดยตรง
- ไม่เหมาะสมกับข้อมูลที่เป็นไบนาリー (มีแค่ 2 ค่า เช่น ใช่/ไม่ใช่)

**sysuse nlsw88.dta, clear  
tab occupation industry**

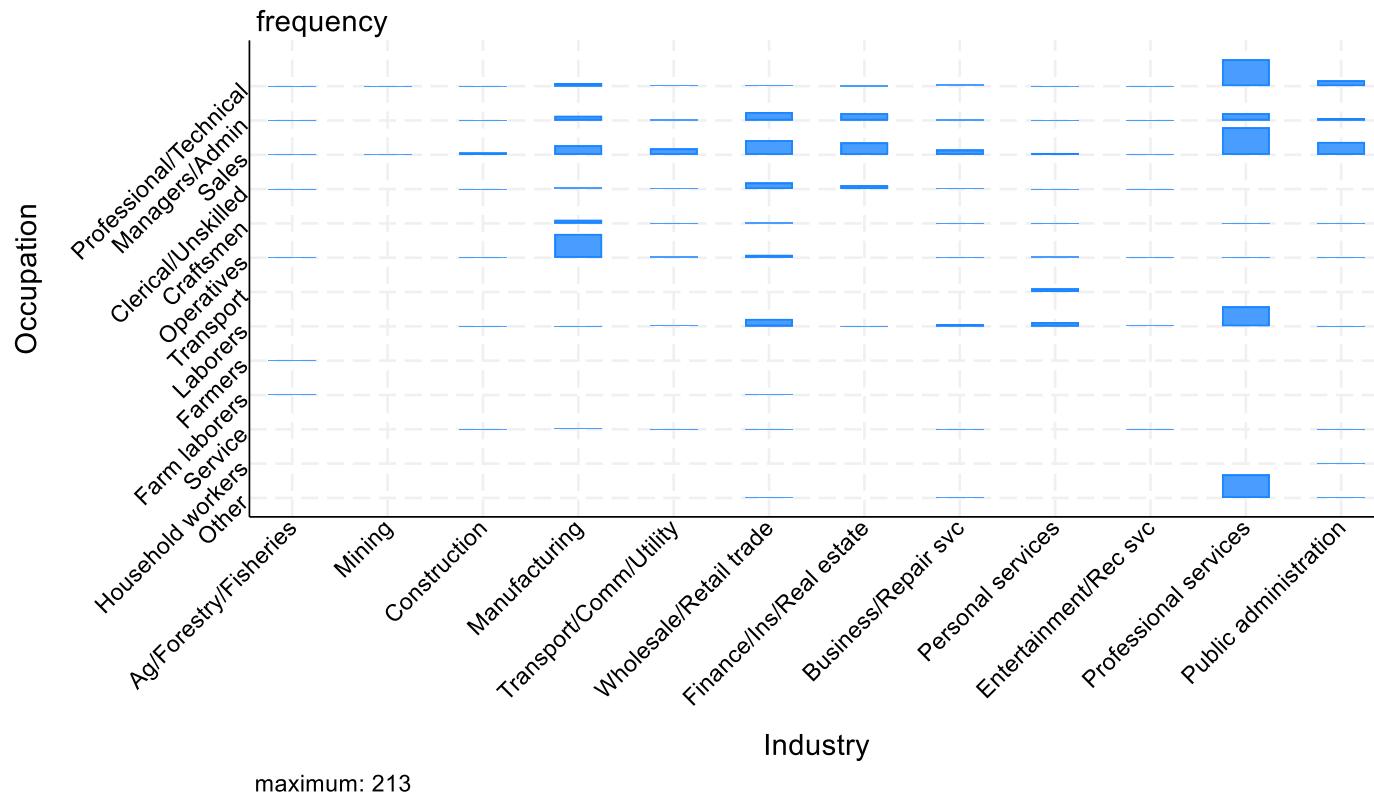
Occupation	Industry										Total
	Ag/Forest	Mining	Construct	Manufactu	Transport	Wholesale	Finance/I	Business/	Personal		
Professional/Technica	1	1	1	24	4	7	8	13	2	316	
Managers/Admin	1	0	5	37	9	65	56	9	5	263	
Sales	6	3	16	73	49	113	95	39	14	725	
Clerical/Unskilled	1	0	1	13	2	50	29	4	1	102	
Craftsmen	0	0	0	29	5	10	0	1	1	53	
Operatives	2	0	2	185	12	23	0	1	12	244	
Transport	0	0	0	0	0	0	0	0	28	28	
Laborers	0	0	1	1	6	58	4	16	33	285	
Farmers	1	0	0	0	0	0	0	0	0	1	
Farm laborers	5	0	0	0	0	2	0	0	0	7	
Service	0	0	3	5	2	2	0	2	0	16	
Household workers	0	0	0	0	0	0	0	0	0	2	
Other	0	0	0	0	0	2	0	1	0	186	
Total	17	4	29	367	89	332	192	86	96	2,228	

Occupation	Industry			Total
	Entertain	Professio	Public ad	
Professional/Technica	2	208	45	316
Managers/Admin	2	56	18	263
Sales	4	213	100	725
Clerical/Unskilled	1	0	0	102
Craftsmen	0	5	2	53
Operatives	1	5	1	244
Transport	0	0	0	28
Laborers	6	155	5	285
Farmers	0	0	0	1
Farm laborers	0	0	0	7
Service	1	0	1	16
Household workers	0	0	2	2
Other	0	182	1	186

```

ssc install tabplot
tabplot occupation industry, ylabel(1(1)13, labsizesmall)
value label angle(45) xlabel(1(1)12, labsizesmall) value label
angle(45)

```



## แผนภูมิบลูน

แผนภูมิบลูน (Balloon Plot) เป็น รูปแบบหนึ่งของการแสดงข้อมูลในตารางสองทางคล้ายกับแผนภูมิตาราง (Table Plot)

จะแสดงข้อมูลด้วยวงกลม (Circles) โดยขนาดของวงกลม จะแสดงถึงค่าความถี่หรือค่าทางสถิติที่เกี่ยวข้อง

- ดูน่ามองกว่าแผนภูมิตаблиц - วงกลมให้ความรู้สึกที่ต่อเนื่องกว่าแท่งกราฟ
- สามารถเปรียบเทียบขนาดของข้อมูลได้ทุกทิศทาง
- เหมาะสมสำหรับข้อมูลที่ต้องการนำเสนออย่างเรียบง่ายและไม่ซับซ้อน

## แผนภูมิแท่งซ้อน (Stacked Bar Chart)

แสดงข้อมูลที่ใช้แท่งกราฟแบบซ้อนกัน เพื่อแสดงสองตัวแปรเชิงหมวดหมู่ (Categorical Variables) พร้อมกัน โดยการแบ่งแท่งหลักออกเป็นแท่งย่อย ซึ่งแต่ละแท่งย่อยแสดงระดับของตัวแปรที่สอง

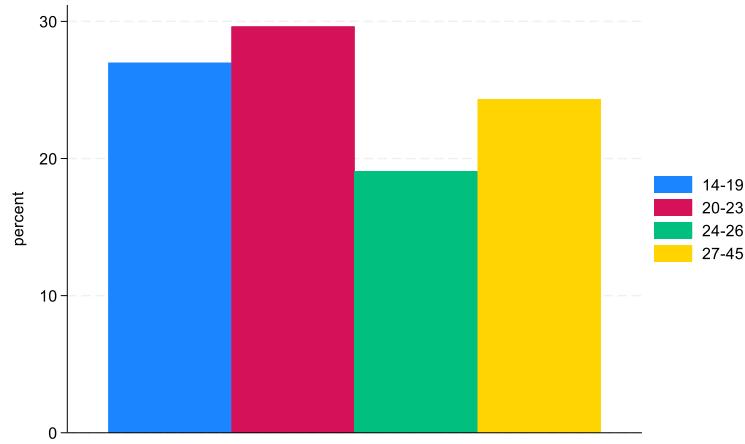
### ข้อดี:

- แสดงความแตกต่างระหว่างสองตัวแปรพร้อมกัน
- ช่วยให้เข้าใจแนวโน้มได้ง่ายขึ้น
- สามารถใช้ได้ทั้งแบบแนวตั้งและแนวนอน → แนวอนเหมาะสมกับข้อมูลที่มีข้อความยาว

### ข้อเสีย:

- อ่านค่าสัดส่วนได้ยากสำหรับหมวดหมู่ที่อยู่ตรงกลางของแท่งซ้อน
- การเปรียบเทียบค่าระหว่างแท่งหลักอาจไม่ชัดเจนเท่ากับ Bar Chart ปกติ
- ถ้ามีหมวดหมู่มากเกินไป แผนภูมิอาจดูยุ่งเหยิงและอ่านยาก

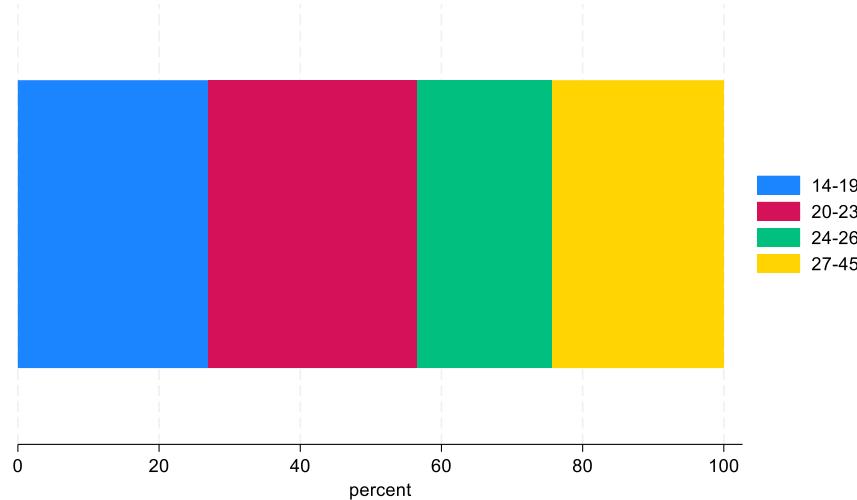
webuse lbw.dta, clear  
graph bar, over(agecats) asyvars



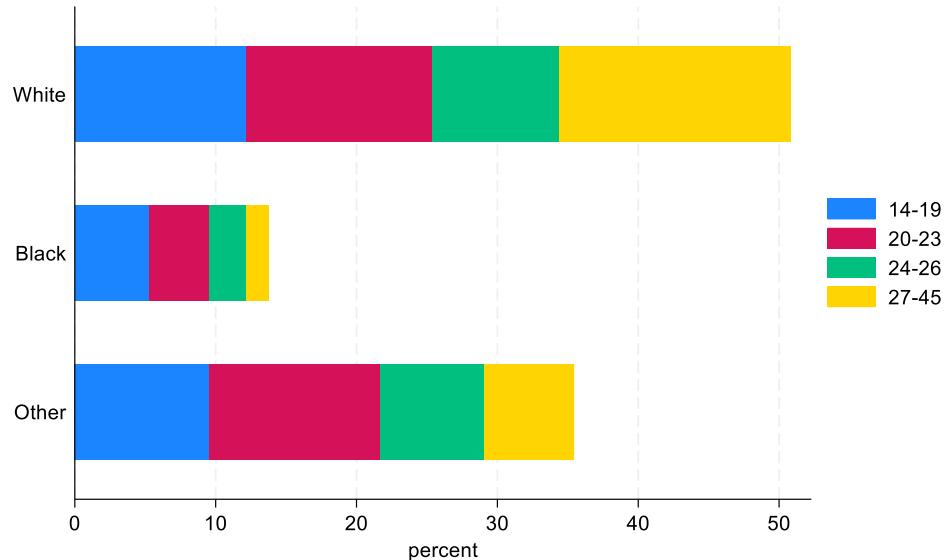
graph bar, over(agecats) stack asyvars



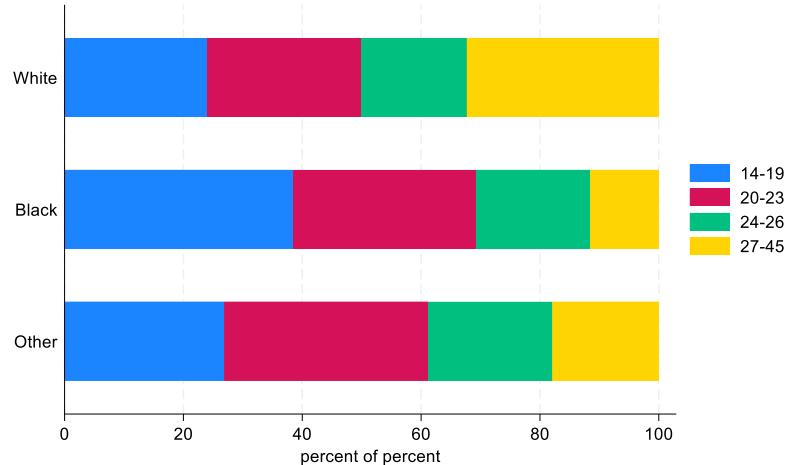
**graph hbar, over(agecats) stack asyvars**



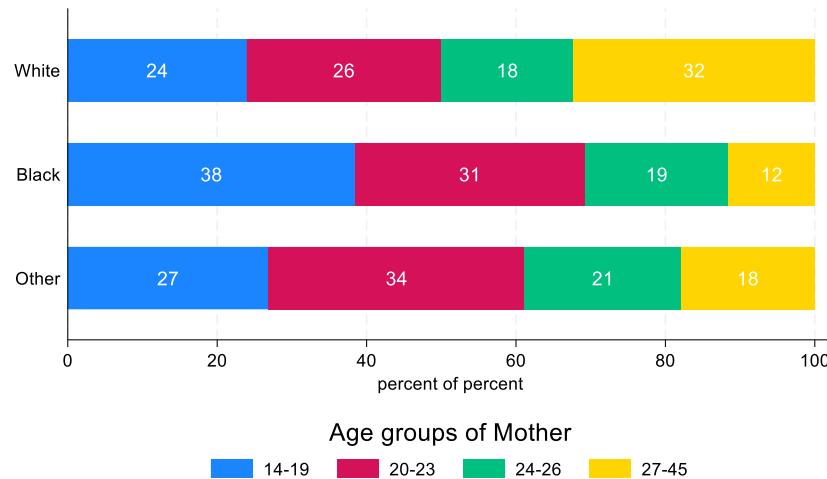
**graph hbar, over(agecats) over(race) stack asyvars**



```
graph hbar, over(agecats) over(race) stack percentage asyvars
```



```
graph hbar, over(agecats) over(race) stack asyvars percentage blabel(bar,  
pos(center) format(%3.0f) size(medium) color(white)) legend(position(6) row(1)  
size(medsmall) title("Age groups of Mother"))
```



## 7. กราฟสำหรับตัวแปรสามตัวหรือมากกว่า (Three or More Variables)

---

ขยายจากแผนภูมิที่มีตัวแปรสองตัวแปร โดยเพิ่มตัวแปรที่สามในแกน 2 มิติ

## แผนภูมิ Contour

แผนภูมิ Contour (Contour Plot) เป็น แผนภูมิแบบ 3 มิติที่แสดงผลใน 2 มิติ

- X และ Y → ตัวแปรเชิงปริมาณที่ใช้พล็อตในแกนแนวตั้งและแนวนอน
- Z → ตัวแปรที่แสดงค่าคงที่ในแต่ละจุดผ่านเส้น Contour

หลักการสำคัญ:

- เส้น Contour แสดงค่าคงที่ของตัวแปรที่สาม (Z) บนแกน X และ Y
- นิยมใช้ใน แผนที่ภูมิประเทศ หรือ การวิเคราะห์แนวโน้มข้อมูล
- ใช้ได้ดีกับข้อมูลที่มีรูปแบบแนวโน้มที่ชัดเจน

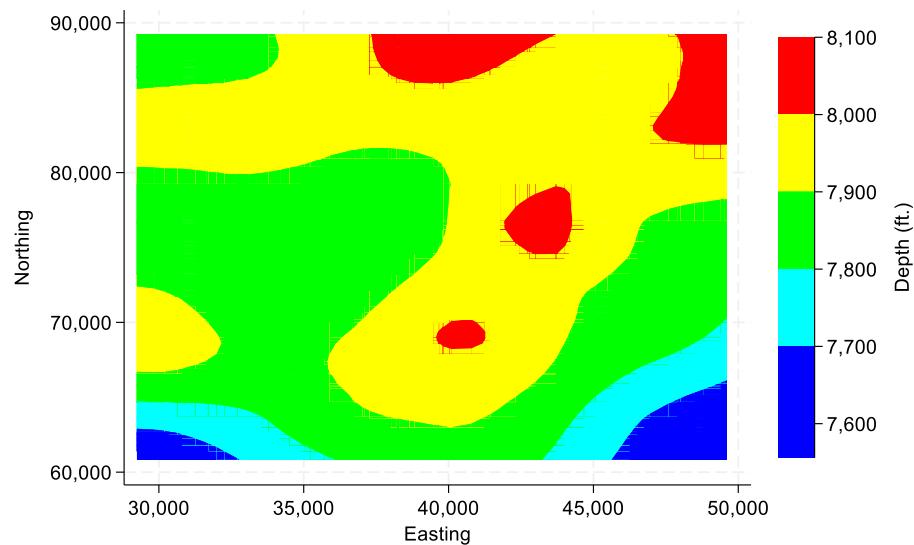
## ข้อดี:

- ช่วยแสดงข้อมูล 3 มิติในรูปแบบ 2 มิติที่เข้าใจง่าย
- เหมาะกับการแสดงแนวโน้มของข้อมูลที่มีความต่อเนื่อง
- สามารถใช้กับข้อมูลพารามิเตอร์ เช่น ผลลัพธ์ของแบบจำลองทางเศรษฐศาสตร์

## ข้อเสีย:

- ต้องใช้ข้อมูลปริมาณมาก → ถ้าข้อมูลน้อยเกินไป จะไม่สามารถสร้าง Contour ที่แม่นยำได้
- ต้องกำหนดจำนวนเส้น Contour อย่างเหมาะสม → หากไปจะซับซ้อน, น้อยไปจะไม่เห็นรายละเอียด
- ใช้ได้กับข้อมูลต่อเนื่องเท่านั้น → ไม่เหมาะสมกับตัวแปรประเภทหมวดหมู่ (Categorical)
- ต้องพึ่งพาการประมาณค่า (Interpolation) → อาจทำให้ค่าที่แสดงออกมาคลาดเคลื่อน

**sysuse sandstone, clear  
tw (contour depth northing easting)**



## แผนภูมิฟองอากาศ (Bubble Plot)

แผนภูมิฟองอากาศ (Bubble Plot) เป็น วิธีที่ช่วยแสดงผลข้อมูลสามมิติ โดยไม่ต้องใช้แกน Z

- X และ Y → เป็นตัวแปรที่พล็อตบนแกนแนวตั้งและแนวนอนตามปกติ
- ขนาดของฟองอากาศ (Bubble Size) → ใช้แทนค่าของตัวแปรที่สาม (Z)
- ขนาดฟองอากาศต้องคำนวณจากพื้นที่ (Area) ไม่ใช่รัศมี (Radius) เพื่อให้สัดส่วนมีความถูกต้อง

ตัวอย่างเช่น เราพิจารณาความสัมพันธ์ระหว่าง GDP ต่อหัว (X) และอายุขัยเฉลี่ย (Y) ของแต่ละประเทศ  
เราต้องการพิจารณาขนาดของประเทศซึ่งกำหนดโดยประชากร (Z)

## ข้อดี:

- เข้าใจง่ายและเหมาะสมสำหรับผู้ชมทั่วไป
- สามารถแสดงผลข้อมูลสามมิติได้ในสองมิติ
- สามารถขยายเป็น 4-5 มิติได้ โดยใช้สีและรูปแบบของเครื่องหมาย
- นิยมใช้ในการนำเสนอข้อมูลสถิติที่ซับซ้อนแบบง่ายๆ

## ข้อเสีย:

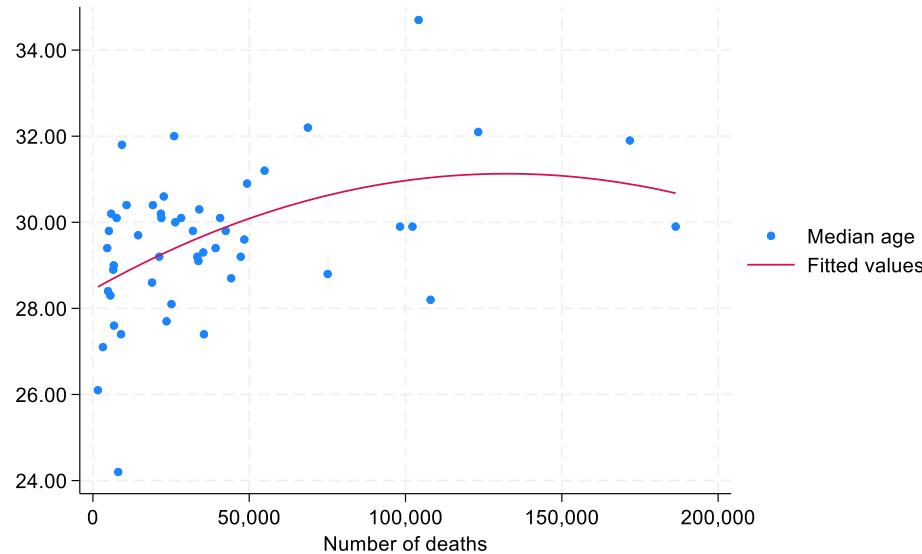
- ยากต่อการอ่านค่าตัวเลขที่แม่นยำ → ขนาดของฟองอากาศอาจทำให้สัดส่วนคลาดเคลื่อน
- เมื่อข้อมูลมากเกินไป ฟองอากาศอาจทับซ้อนกันจนอ่านยาก
- ไม่สามารถแสดงค่าลับได้โดยตรง → ต้องใช้สีหรือสัญลักษณ์อื่นแทน
- ความเข้าใจผิดเกี่ยวกับความสัมพันธ์ของตัวแปร → ควรเพิ่มเส้นแนวโน้ม (Trend Line)

## การปรับแต่งแผนภูมิพองอากาศ

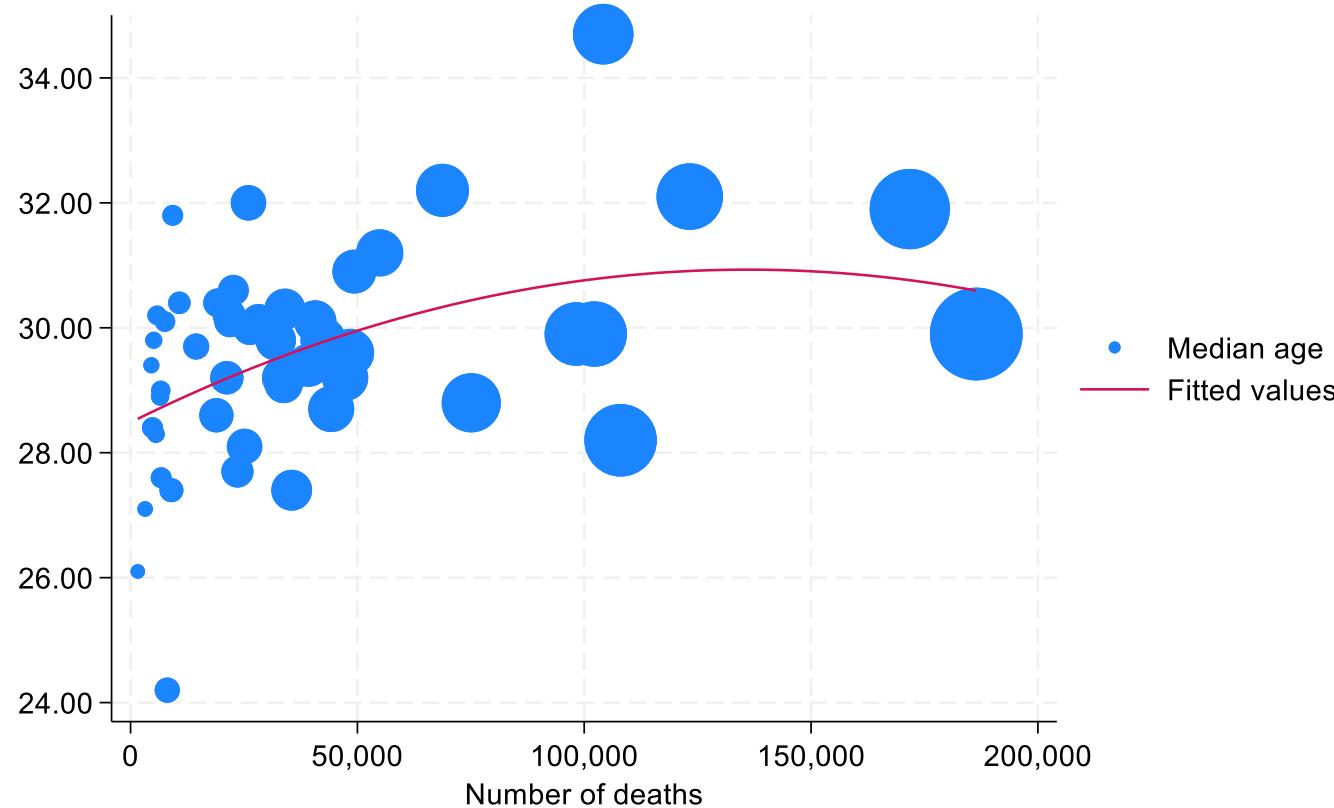
- การเพิ่มป้ายกำกับ (Marker labels)
- การเปลี่ยนรูปร่างของเครื่องหมาย เช่น อาจจะเพิ่มทวีปเข้าไป เป็นมิติที่ 4
- การใช้สีแทนค่าของตัวแปร
- การเพิ่มเส้นแนวโน้ม

```
sysuse census, clear
```

```
tw (scatter medage death) (qfit medage death)
```



```
tw (scatter medage death [fweight = pop]) (qfit medage death  
[fweight = pop])
```



```

tw (scatter medage death if region == 1 [fweight = pop], mcolor(red) msymbol(O)) (scatter
medage death if region == 2 [fweight = pop], mcolor(blue) msymbol(D)) (scatter medage death
if region == 3 [fweight = pop], mcolor(green) msymbol(S)) (scatter medage death if region
== 4 [fweight = pop], mcolor(orange) msymbol(T)) (qfit medage death [fweight = pop]) ,
legend(position(6) row(1) order(1 "NE" 2 "N Cntrl" 3 "South" 4 "West" 5 "Fit"))

```

