

ข้อมูลห้ศจรรยและะถึนที่อยุ่

สนุกกับข้อมูล เรียนรู้แนวคิถพื้นฐาน
ทดลองตั้งโจทยจกข้อมูลทีหาได้
สัฒผัศควมสฬสในการนำข้อมูลไปใช้

จิตรทัศน์ ผักเจริญผล – ภาควิชาวิศวกรรมคอมพิวเตอร์ ม.เกษตรศาสตร์

Image by Gerd Altmann

From <https://pixabay.com/illustrations/web-connection-data-wave-binary-4809584/>

กิจกรรม 1

- แบ่งกลุ่ม กลุ่มละ 10 คน
- เก็บข้อมูลของสมาชิกในกลุ่ม:
 - ส่วนสูง
 - ขนาดรอบหัว
 - ขนาดรอบข้อมือ
 - ขนาดเท้า
 - อาชีพ

ข้อมูลแบบมีโครงสร้าง

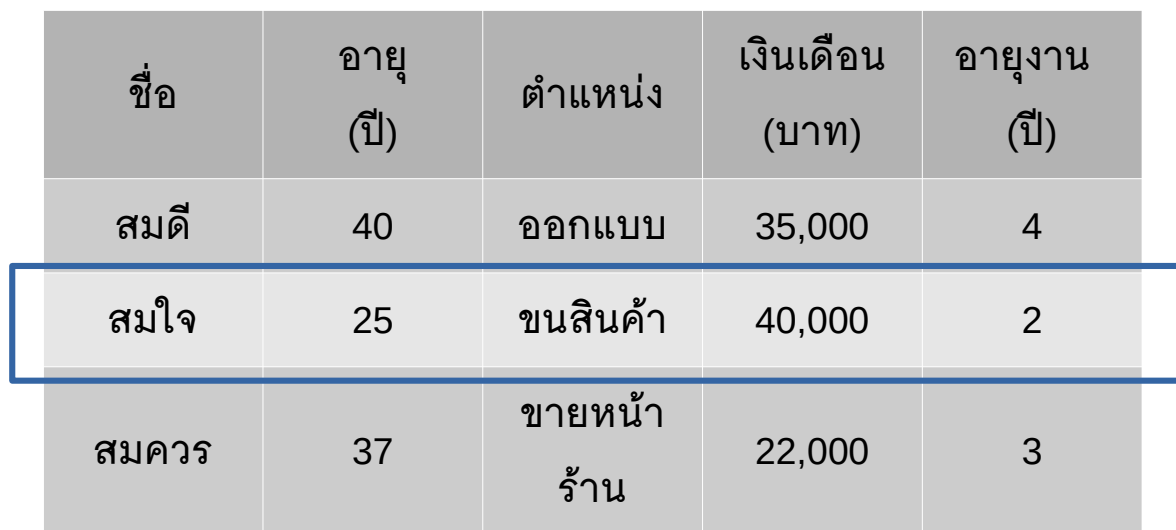
- รูปแบบของ “ข้อมูล” ที่เราหาได้มักอยู่ในรูปของตาราง

ชื่อ	อายุ (ปี)	ตำแหน่ง	เงินเดือน (บาท)	อายุงาน (ปี)
สมดี	40	ออกแบบ	35,000	4
สมใจ	25	ขนส่งสินค้า	40,000	2
สมควร	37	ขายหน้าร้าน	22,000	3

ศัพท์แสง (หลายชื่อเรียก)

คุณลักษณะ (attribute, feature)

ตัวแปร (variable)



ชื่อ	อายุ (ปี)	ตำแหน่ง	เงินเดือน (บาท)	อายุงาน (ปี)
สมดี	40	ออกแบบ	35,000	4
สมใจ	25	ขนส่งสินค้า	40,000	2
สมควร	37	ขายหน้า ร้าน	22,000	3

ตัวอย่าง (example)

กรณี (case)

เรคอร์ด

จำนวนคุณลักษณะ = มิติ

ตัวอย่างข้อมูลเชิงสถิติ (1)

	A	B	C	D	E	F	G	H	I	J	K	L
1	จำนวนคนจน (ด้านรายจ่าย) จำแนกเป็นรายภาค และจังหวัด พ.ศ. 2552 - 2561											
2	หน่วย: พันคน											
3												
4	ภาค	จังหวัด	2552	2553	2554	2555	2556	2557	2558	2559	2560	2561
5	ทั่วราชอาณาจักร	ทั่วราชอาณาจักร	11,623.9	10,800.7	8,751.9	8,402.1	7,305.1	7,057.4	4,847.2	5,810.1	5,324.8	6,682.5
6	กรุงเทพมหานคร	กรุงเทพมหานคร	190.5	186.7	647.9	161.5	90.4	140.6	173.7	118.0	98.3	123.6
7	ภาคกลาง	ภาคกลาง	1,971.8	1,957.9	1,905.1	1,291.5	1,014.9	941.5	827.3	1,010.4	884.0	1,374.9
8		สมุทรปราการ	72.3	27.7	34.6	16.8	20.1	20.2	7.5	6.0	0.8	7.0
9		นนทบุรี	27.3	17.8	16.4	9.1	11.9	2.8	9.2	10.6	11.5	11.8
10		ปทุมธานี	58.5	12.5	49.2	2.2	1.6	1.6	3.7	1.6	0.0	23.6
11		พระนครศรีอยุธยา	60.5	48.5	59.8	33.6	14.0	23.8	17.7	29.2	21.6	86.2
12		อ่างทอง	33.9	42.6	36.6	24.6	22.6	35.5	37.5	37.7	42.8	60.8
13		ลพบุรี	156.1	195.4	167.3	94.4	68.1	95.5	60.7	81.7	80.6	122.5
14		สิงห์บุรี	29.9	44.8	19.0	17.9	5.8	12.3	12.4	20.3	20.6	38.6
15		ชัยนาท	52.9	51.0	30.7	67.3	41.0	70.9	61.4	88.9	82.4	87.4
16		สระบุรี	90.1	97.8	59.3	32.6	43.8	43.2	35.7	25.5	11.2	50.5
17		ชลบุรี	22.8	32.1	38.3	1.2	1.3	6.4	10.0	0.5	35.5	5.7
18		ระยอง	47.9	29.4	30.1	17.9	24.1	2.8	4.7	19.3	16.9	39.7
19		จันทบุรี	49.9	50.3	58.5	69.8	66.4	45.0	19.4	19.8	33.3	29.9
20		ตราด	61.0	65.9	50.2	52.5	45.2	29.0	30.3	37.9	22.3	41.2

Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector_08_11_TH_.xlsx
 From: <http://statbbi.nso.go.th/staticreport/page/sector/th/08.aspx>

ตัวอย่าง (2)

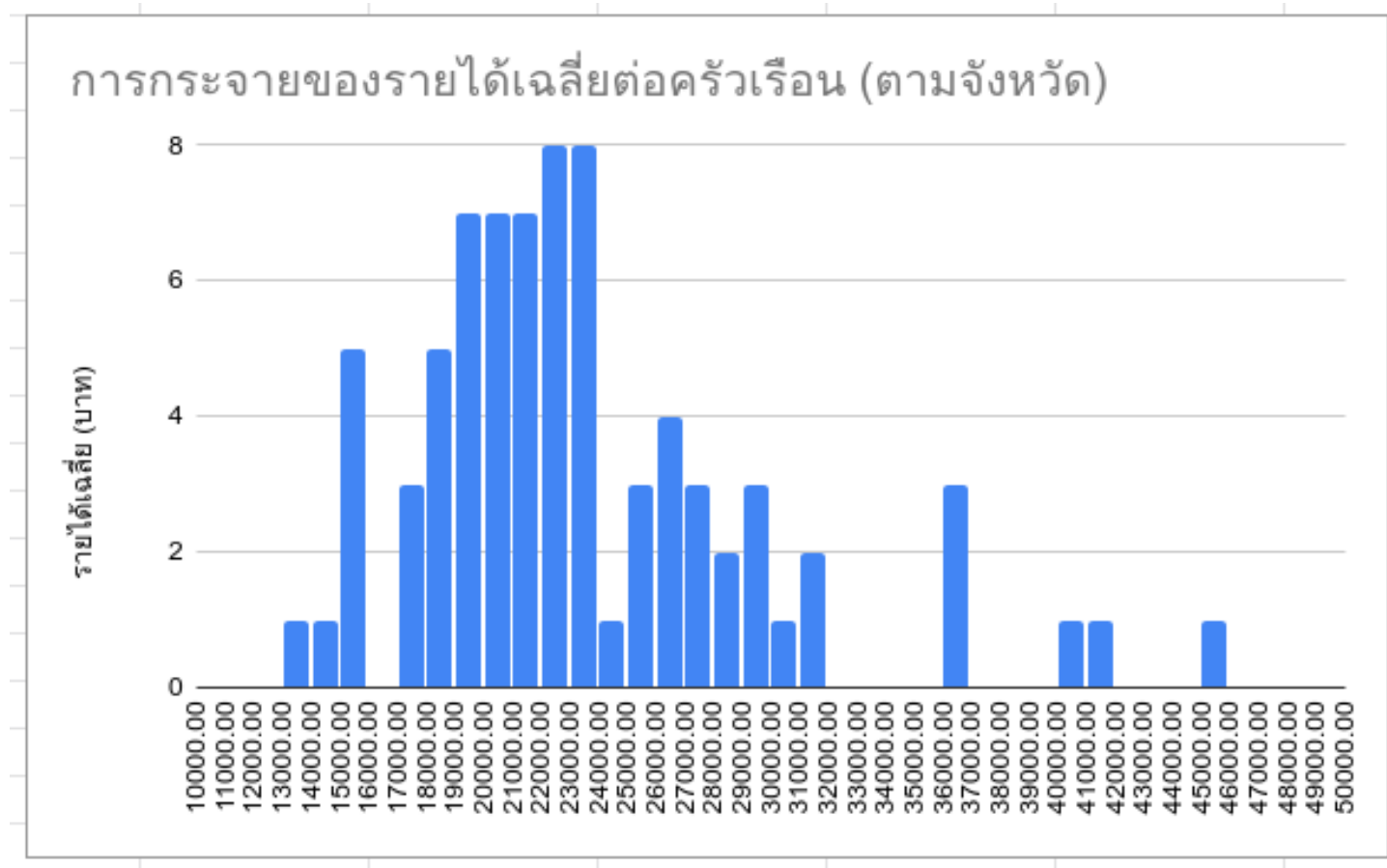
1	หนี้สินเฉลี่ยต่อครัวเรือน จำแนกตามวัตถุประสงค์ของการกู้ยืม เป็นรายจังหวัด พ.ศ. 2545 - 2560						
2	หน่วย: บาท						
3							
4	ภาค	จังหวัด	วัตถุประสงค์ของการกู้ยืม	2545	2547	2549	2550
5	กรุงเทพมหานคร	กรุงเทพมหานคร	หนี้สินทั้งสิ้น	162,199.23	160,775.93	158,059.00	155,396.00
6			เพื่อใช้จ่ายในครัวเรือน	117,724.88	37,388.46	31,693.00	38,184.00
7			เพื่อใช้ทำธุรกิจที่ไม่ใช่การเกษตร	44,118.82	31,081.73	48,960.00	27,517.00
8			เพื่อใช้ทำการเกษตร	85.69	445.82	1,189.00	3,298.00
9			เพื่อใช้ในการศึกษา	0	0	4,362.00	4,953.00
10			เพื่อใช้ซื้อ/เช่าซื้อบ้านและที่ดิน	0	86,207.61	68,779.00	67,457.00
11			อื่น ๆ	269.82	5,652.28	3,076.00	13,986.00
12	ภาคกลาง	สมุทรปราการ	หนี้สินทั้งสิ้น	84,345.14	78,750.16	95,975.00	85,904.00
13			เพื่อใช้จ่ายในครัวเรือน	71,284.00	23,850.24	21,393.00	31,960.00
14			เพื่อใช้ทำธุรกิจที่ไม่ใช่การเกษตร	11,070.68	9,221.45	5,249.00	3,570.00
15			เพื่อใช้ทำการเกษตร	1,969.19	4,083.22	1,353.00	1,054.00
16			เพื่อใช้ในการศึกษา	0	0	1,292.00	2,729.00
17			เพื่อใช้ซื้อ/เช่าซื้อบ้านและที่ดิน	0	40,699.08	60,788.00	46,590.00
18			อื่น ๆ	21.26	896.15	5,899.00	
19		นนทบุรี	หนี้สินทั้งสิ้น	145,908.08	203,040.52	179,007.00	196,895.00
20			เพื่อใช้จ่ายในครัวเรือน	123,235.70	57,416.23	46,911.00	56,804.00
21			เพื่อใช้ทำธุรกิจที่ไม่ใช่การเกษตร	11,070.68	9,221.45	5,249.00	3,570.00

ที่มาของข้อมูล

- สอบถาม/เก็บจากกลุ่มเป้าหมาย
 - เก็บทั้งหมด
 - สุ่มเก็บบางส่วน (sample)
 - ความคลาดเคลื่อน
 - อคติและความลำเอียง (bias)
 - ข้อมูลตกหล่น

ข้อมูลบอกอะไรกับเรา

- ลักษณะการกระจาย

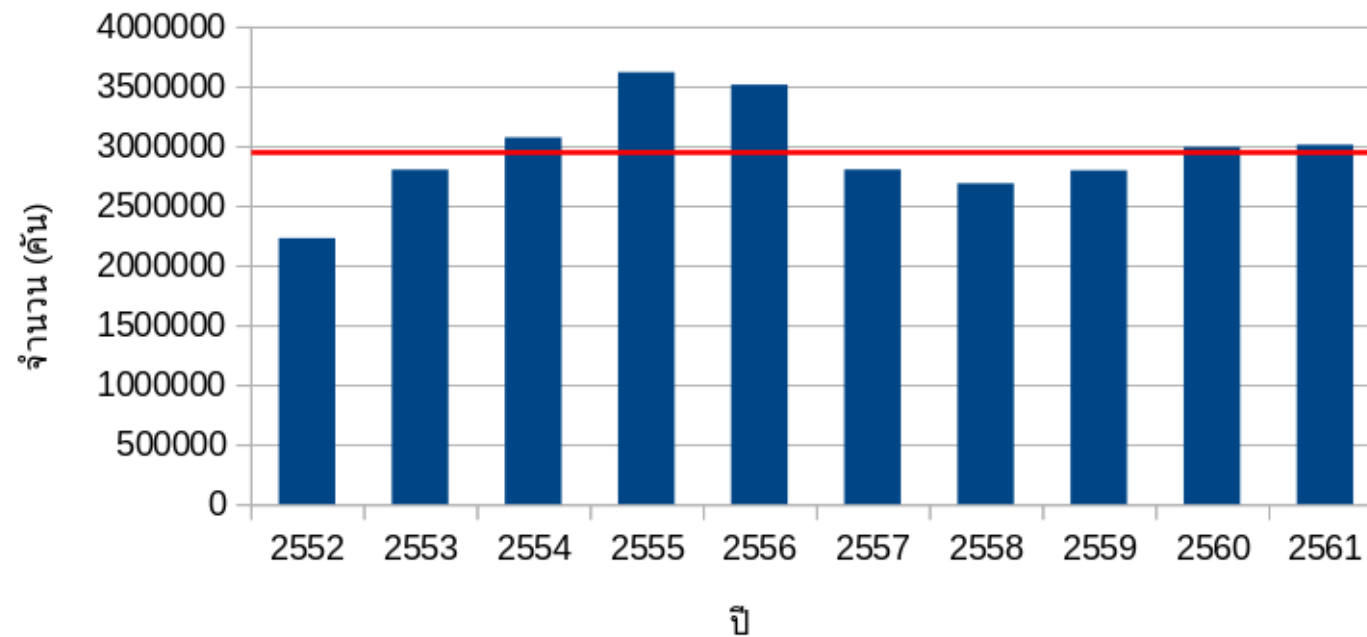


ข้อมูลบอกอะไรกับเรา

- “ค่ากลาง” (1)

จำนวนรถยนต์จดทะเบียนใหม่

ตามกฎหมายว่าด้วยรถยนต์

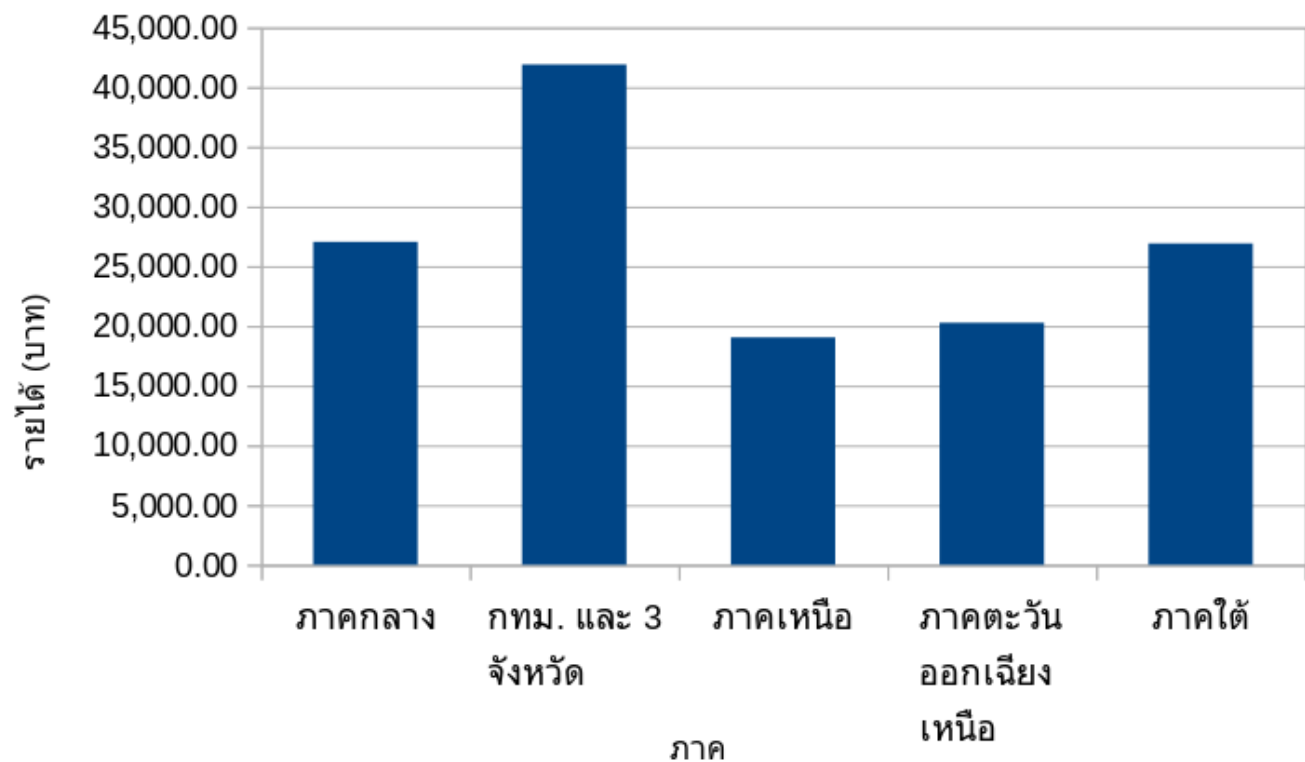


ข้อมูลบอกอะไรกับเรา

• “ค่ากลาง” (2)

	A	B	C
5	ทั่วราชอาณาจักร	ทั่วราชอาณาจักร	26,946.43
6	กรุงเทพมหานคร และ 3	ดอนพิเศษ	41,897.06
7	จังหวัด	กรุงเทพมหานคร	45,707.31
8		สมุทรปราการ	28,711.77
9		นนทบุรี	40,860.88
10		ปทุมธานี	41,483.71
11	ภาคกลาง (ไม่รวม กทม. และ 3 จังหวัดรอบ กทม.)	กทม. และ 3 จังหวัด	27,041.89
12		พระนครศรีอยุธยา	28,777.58
13		อ่างทอง	26,484.17
14		ลพบุรี	19,795.67
15		สิงห์บุรี	26,865.56
16		ชัยนาท	26,385.66
17		สระบุรี	35,378.49
18		ชลบุรี	27,665.39
19		ระยอง	27,797.79
20		จันทบุรี	32,893.60
21		ตราด	27,796.68
22		ฉะเชิงเทรา	26,061.85
23		ปราจีนบุรี	22,952.89
24		นครนายก	25,119.93
25		สระแก้ว	22,115.02
26		ราชบุรี	33,622.28
27		กาญจนบุรี	20,564.74
28		สุพรรณบุรี	20,440.57
29		นครปฐม	32,760.69
30		สมุทรสาคร	25,446.01
31		สมุทรสงคราม	30,182.85

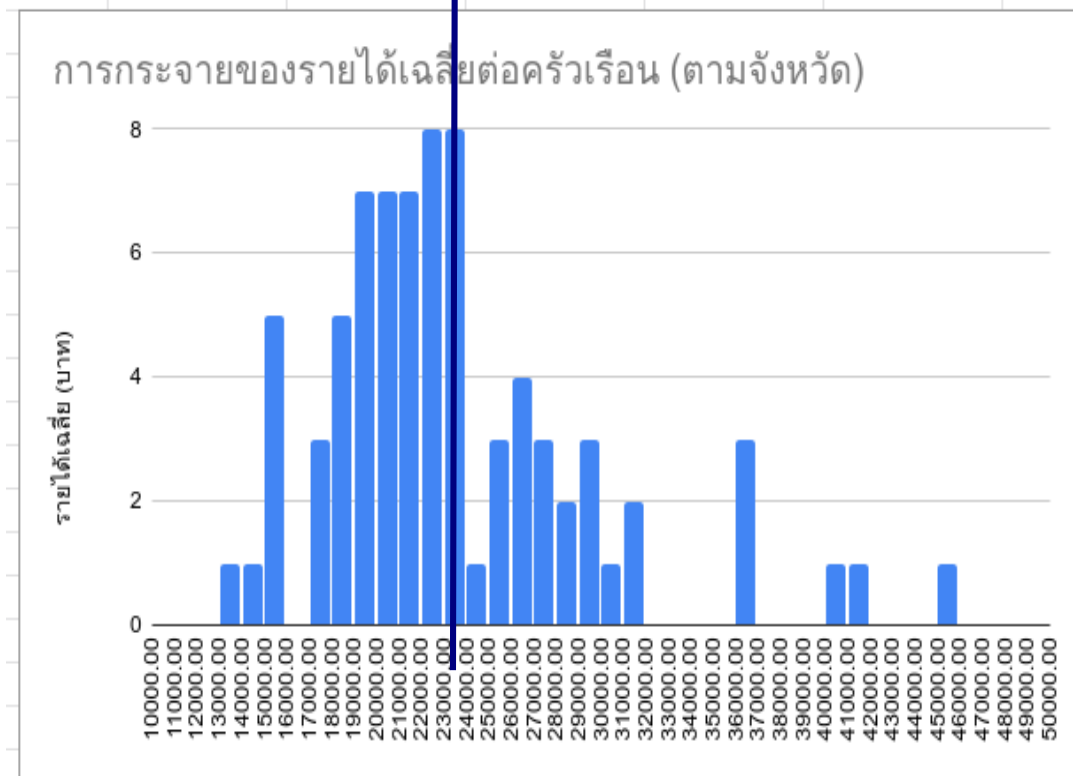
รายได้เฉลี่ย แยกตามภาค



ข้อควรระวัง - ค่ากลาง

A	B	C	D	E
จังหวัด	Province	ภาค	Region	year-2558-2015
กรุงเทพมหานคร	Bangkok	ตอนพิเศษ	Greater Bangkok	45571.7
สมุทรปราการ	Samut Prakan	ตอนพิเศษ	Greater Bangkok	25457.2
นนทบุรี	Nonthaburi	ตอนพิเศษ	Greater Bangkok	36884
ปทุมธานี	Pathum Thani	ตอนพิเศษ	Greater Bangkok	41056.9
พระนครศรีอยุธยา	Phra Nakhon Si	ภาคกลาง	Central Region	28379.4
อ่างทอง	Ang Thong	ภาคกลาง	Central Region	23350.6
ลพบุรี	Lop Buri	ภาคกลาง	Central Region	22955.1
สิงห์บุรี	Sing Buri	ภาคกลาง	Central Region	26112.2
ชัยนาท	Chai Nat	ภาคกลาง	Central Region	22058.9
สระบุรี	Saraburi	ภาคกลาง	Central Region	29413.4
ชลบุรี	Chon Buri	ภาคกลาง	Central Region	27256.7
ระยอง	Rayong	ภาคกลาง	Central Region	30314.8
จันทบุรี	Chanthaburi	ภาคกลาง	Central Region	36023.5
ตราด	Trat	ภาคกลาง	Central Region	25332.9
ฉะเชิงเทรา	Chachoengsao	ภาคกลาง	Central Region	27554.9
ปราจีนบุรี	Prachin Buri	ภาคกลาง	Central Region	24165.6
นครนายก	Nakhon Nayok	ภาคกลาง	Central Region	23555
สระแก้ว	Sa Kaeo	ภาคกลาง	Central Region	26953.4
ราชบุรี	Ratchaburi	ภาคกลาง	Central Region	19590

ค่าเฉลี่ย = 23,542.27 บาท

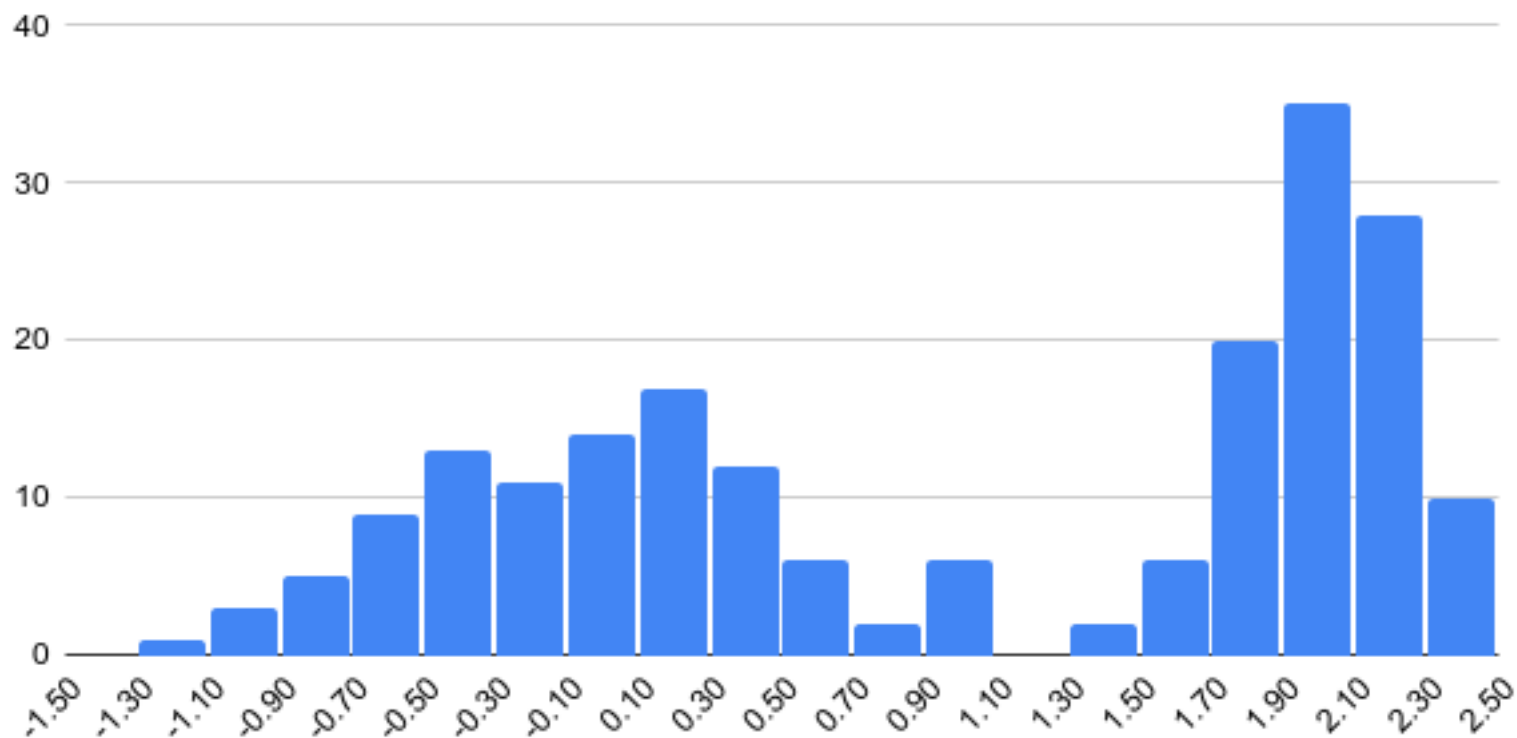


?? เป็นค่าเฉลี่ยของอะไร ??

ข้อมูลบอกอะไรกับเรา

- กลุ่มก้อน (ผ่านทางลักษณะการกระจาย) [1 มิติ]

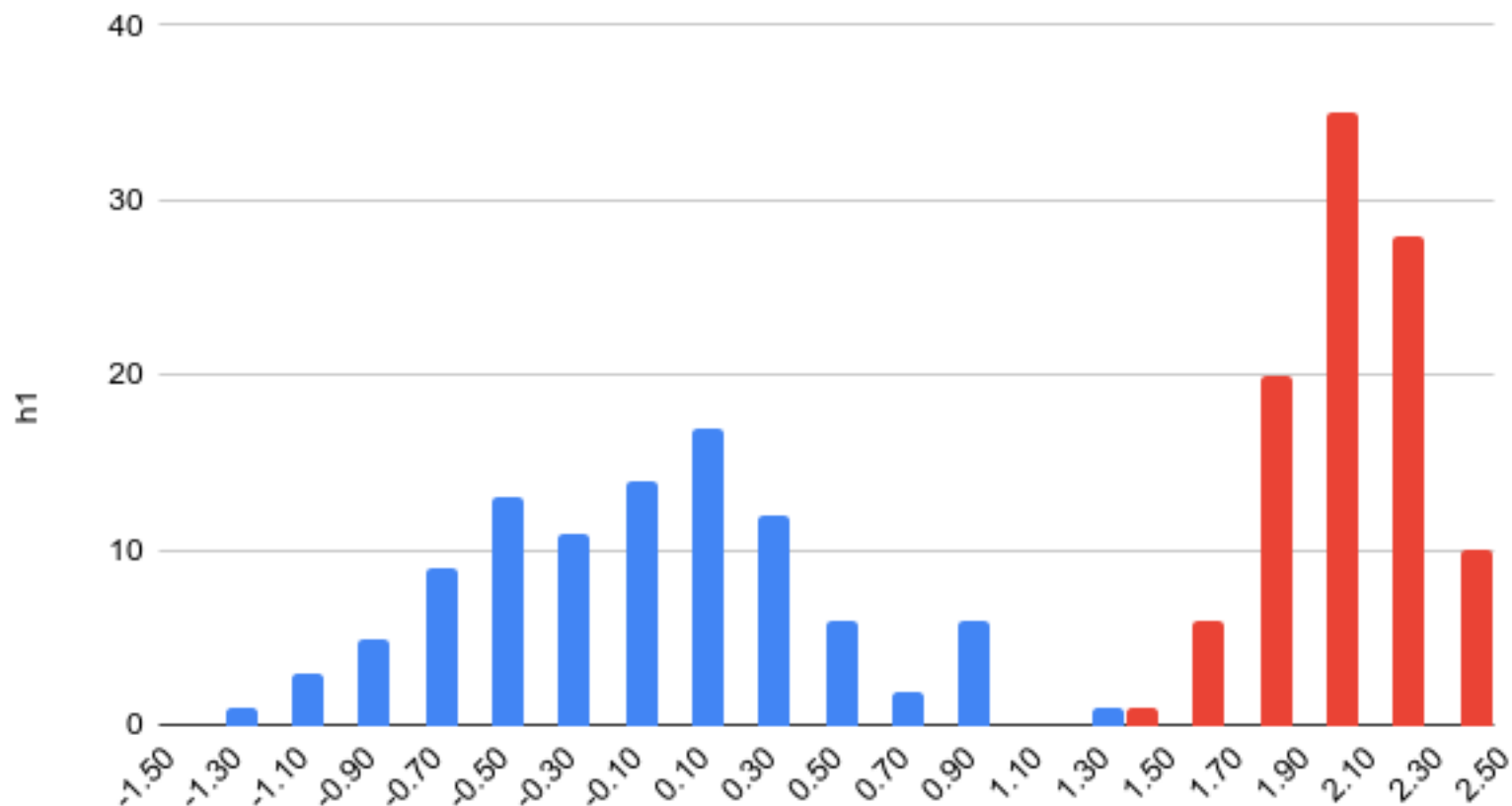
Histogram of combined distributions



ข้อมูลบอกอะไรกับเรา

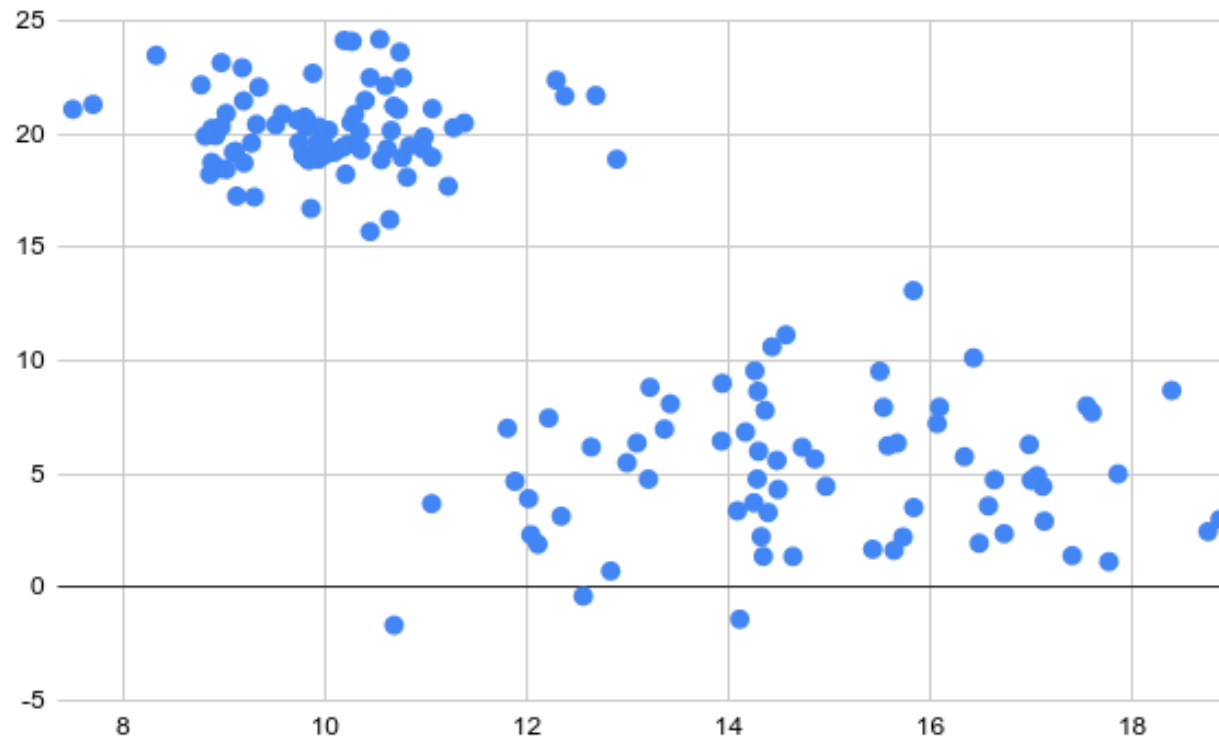
- กลุ่มก้อน (ผ่านทางลักษณะการกระจาย) [1 มิติ]

Histogram of two distribution



ข้อมูลบอกอะไรกับเรา

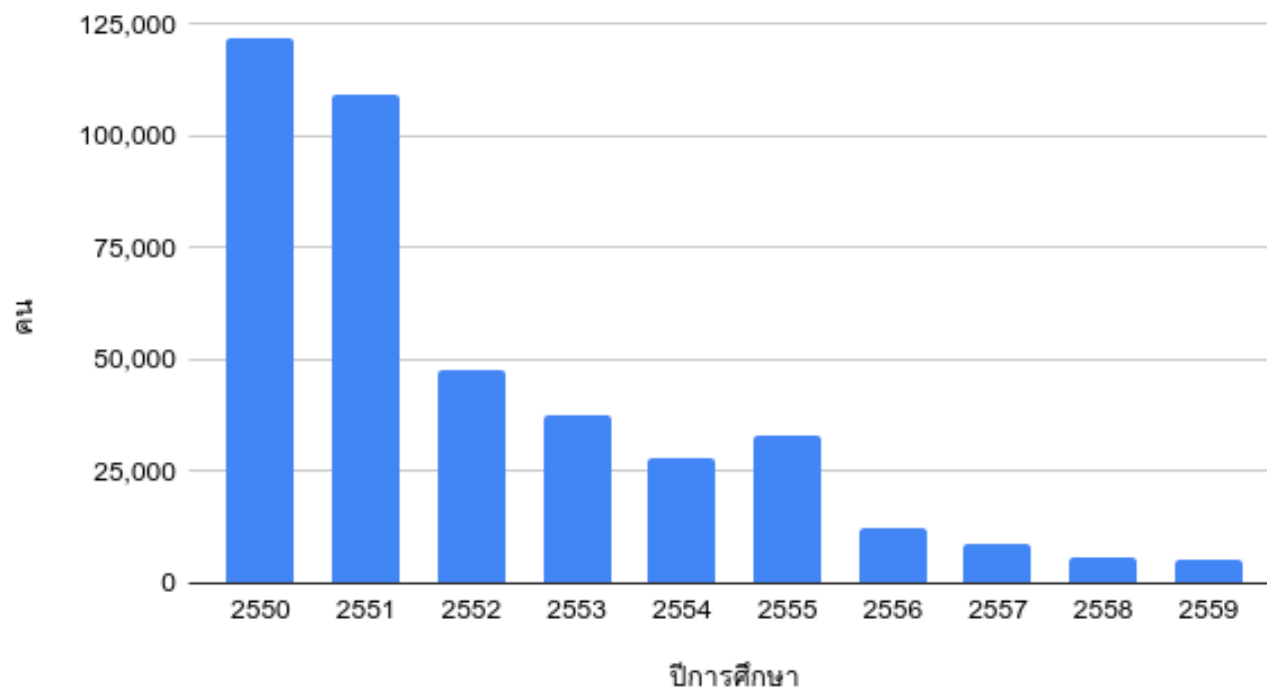
- กลุ่มก้อน (ผ่านทางลักษณะการกระจาย) [2 มิติ]



ข้อมูลบอกอะไรกับเรา

- แนวโน้ม

จำนวนนักเรียนออกกลางคัน (สพฐ.)

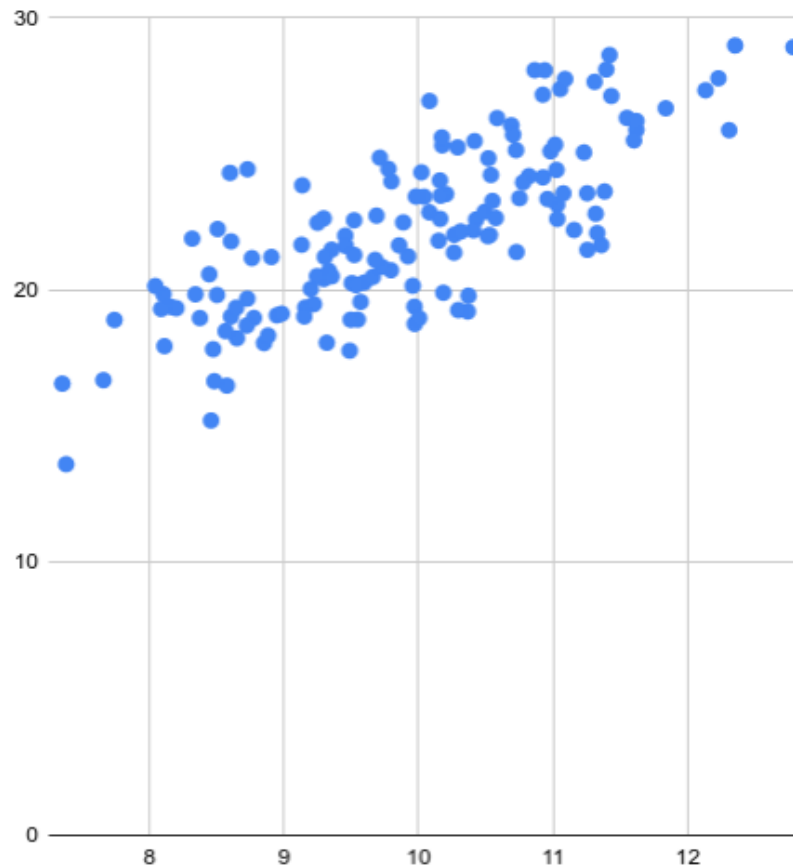


Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector_03_20_TH_.xlsx

From: <http://statbbi.nso.go.th/staticreport/page/sector/th/03.aspx>

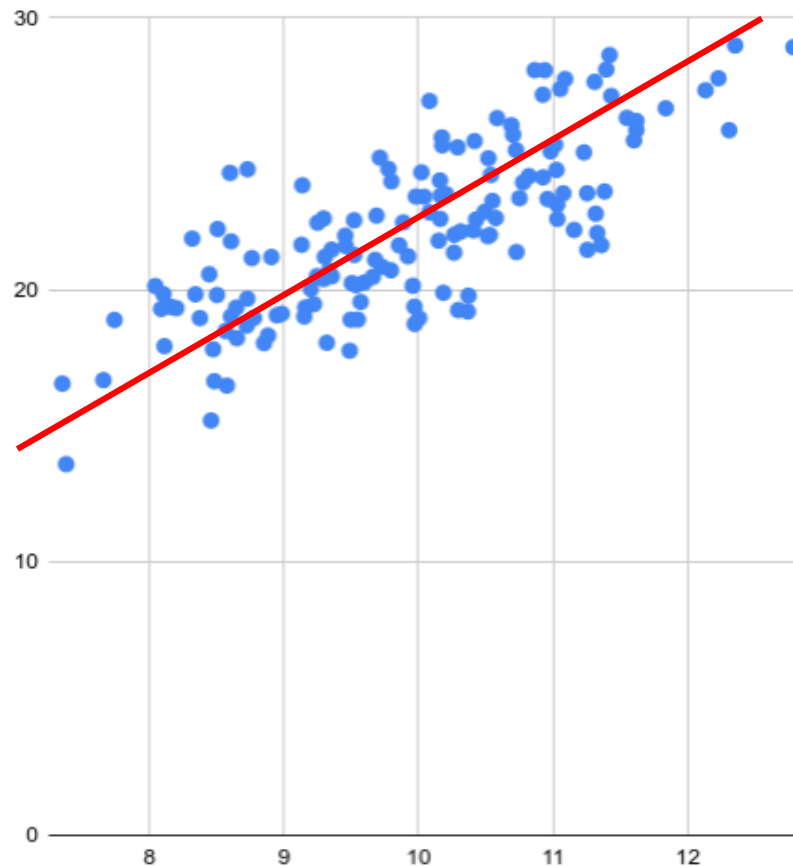
ข้อมูลบอกอะไรกับเรา

- ความสัมพันธ์



ข้อมูลบอกอะไรกับเรา

- การทำนาย (prediction)



CRAAP test

- Currency
- Relevance
- Authority
- Accuracy
- Purpose

Source: https://en.wikipedia.org/wiki/CRAAP_test

CRAAP test was developed by Sarah Blakeslee at California State University, Chico.

การเชื่อมโยงข้อมูล

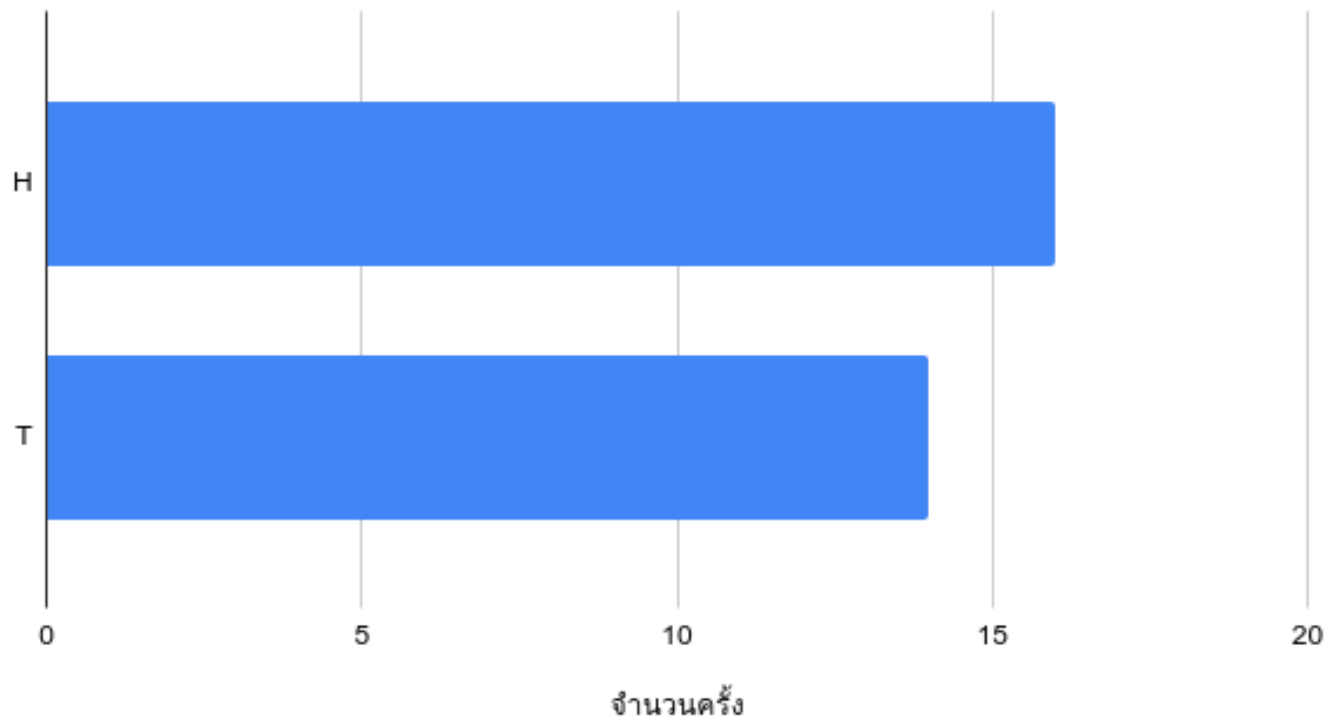
	A	B	C	D	E	F
1	รายได้โดยเฉลี่ยต่อเดือนต่อครัวเรือน เป็นรายภาค และจังหวัด พ.ศ. 2545 - 2560					
2	หน่วย: บาท					
3						
4	ภาค	จังหวัด	2545	2547	2549	2550
5	ทั่วราชอาณาจักร	ทั่วราชอาณาจักร	13,736.00	14,963.40	17,787.00	18,660.00
6	กรุงเทพมหานคร และ 3 จังหวัด	ดอนพิเศษ	28,238.66	28,135.36	33,088.00	35,007.00
7		กรุงเทพมหานคร	29,589.08	29,842.68	36,658.00	39,020.00
8		สมุทรปราการ	19,679.94	19,946.54	20,382.00	21,302.00
9		นนทบุรี	29,119.13	26,657.97	31,152.00	32,743.00
10		ปทุมธานี	22,838.27	21,529.67	25,143.00	26,107.00
11	ภาคกลาง (ไม่รวม กทม. และ 3 จังหวัดรอบ กทม.)	กทม. และ 3 จังหวัด	14,128.00	16,354.53	19,279.00	18,932.00
12		พระนครศรีอยุธยา	13,319.41	14,979.78	19,676.00	21,676.00
13		อ่างทอง	11,652.92	12,854.76	18,300.00	17,704.00
14		ลพบุรี	11,010.41	15,003.26	19,935.00	16,852.00
15		สิงห์บุรี	14,434.26	14,787.81	18,510.00	20,558.00
16		ชัยนาท	11,118.77	13,383.26	13,058.00	13,995.00
17		สระบุรี	14,676.71	18,742.17	20,468.00	22,363.00
18		ชลบุรี	16,797.01	22,286.23	21,869.00	22,260.00
19		ระยอง	13,254.82	18,961.99	20,018.00	25,090.00
20		จันทบุรี	15,958.59	15,896.75	20,606.00	18,866.00
21		ตราด	13,821.60	14,080.46	19,230.00	16,664.00
22		ฉะเชิงเทรา	14,828.66	16,937.66	16,770.00	20,665.00
23		ปราจีนบุรี	12,720.00	15,031.78	16,031.00	18,263.00
24		นครนายก	10,721.74	13,096.37	14,104.00	15,983.00
25		สระแก้ว	9,950.86	10,777.21	11,577.00	13,593.00
26		ราชบุรี	14,692.25	19,866.71	23,202.00	17,576.00

	A	B	C	D	E	F
1	ค่าใช้จ่ายโดยเฉลี่ยต่อเดือนต่อครัวเรือน เป็นรายภาค และจังหวัด พ.ศ. 2552 - 2561					
2	หน่วย: บาท					
3						
4	ภาค	จังหวัด	2552	2553	2554	2555
5	ทั่วราชอาณาจักร	ทั่วราชอาณาจักร	16,205.00	16,819.00	17,403.00	18,766.00
6	กรุงเทพมหานคร และ 3 จังหวัด	กรุงเทพมหานคร และ 3 จังหวัด	27,988.00	28,055.00	27,566.00	31,971.00
7		กรุงเทพมหานคร	30,334.00	29,647.00	29,969.20	33,956.90
8		สมุทรปราการ	18,450.00	19,973.00	20,158.20	25,860.80
9		นนทบุรี	29,495.00	32,423.00	29,099.30	28,731.20
10		ปทุมธานี	22,477.00	23,118.00	18,993.20	30,668.80
11	ภาคกลาง (ไม่รวม กทม. และ 3 จังหวัดรอบ กทม.)	ภาคกลาง (ไม่รวม กทม. และ 3 จังหวัดรอบ กทม.)	17,107.00	16,894.00	16,954.00	19,762.00
12		พระนครศรีอยุธยา	20,717.00	18,789.00	17,534.30	25,215.80
13		อ่างทอง	20,773.00	17,960.00	16,591.20	21,273.20
14		ลพบุรี	16,764.00	14,433.00	14,875.40	17,356.90
15		สิงห์บุรี	17,787.00	15,643.00	19,207.10	23,474.60
16		ชัยนาท	14,272.00	16,110.00	17,391.90	15,535.50
17		สระบุรี	17,378.00	17,228.00	19,377.00	21,412.60
18		ชลบุรี	21,312.00	21,123.00	20,572.80	25,499.00
19		ระยอง	18,566.00	17,689.00	18,546.30	21,023.90
20		จันทบุรี	17,290.00	19,828.00	18,861.10	19,594.70
21		ตราด	13,365.00	13,439.00	17,050.80	15,662.70
22		ฉะเชิงเทรา	19,009.00	18,295.00	17,958.40	23,079.50
23		ปราจีนบุรี	18,472.00	18,565.00	22,883.20	19,733.70
24		นครนายก	13,575.00	12,889.00	13,806.90	16,459.50
25		สระแก้ว	12,396.00	12,697.00	15,129.80	19,531.00

ข้อมูลอาจไม่บอกอะไรกับเรา

- (แต่เราเข้าใจผิด – เรียนเพิ่มวิชา STAT101)

การกระจายของจำนวนครั้งที่ได้หัวและได้ก้อย



ความไม่ปกติของข้อมูล

- ข้อมูลหาย
- ข้อมูลผิด
- ข้อมูลขอบ (outliers)

Bias ในข้อมูล

- Selection bias
- Exclusion bias
- Observer bias

กิจกรรม 2

- จากกลุ่มเดิม แบ่งกลุ่มย่อย
 - plot จุดตัวอย่างลงบนกราฟโดยเลือก **attribute** สองค่า
 - ทำหลาย ๆ แบบ (กลุ่มย่อยละแบบ)
 - พิจารณาแนวโน้มความสัมพันธ์ระหว่าง **attribute**

กระบวนการข้อมูล

- ตั้งคำถาม
- เก็บข้อมูล / เก็บกวาดข้อมูล
- ศึกษาข้อมูล / วิเคราะห์ข้อมูล
- นำเสนอ
- (กลับไปข้อแรก)

ประเภทของคุณสมบัติ (หรือตัวแปร)

- ค่าตัวเลขแบบต่อเนื่อง
- ค่าตัวเลขแบบไม่ต่อเนื่อง
- ประเภท
 - ประเภทที่มีประเภทย่อย
- ช่วง

ข้อมูลแบบประเภท (ตัวอย่าง)

1	จำนวนผู้ป่วยใน (หลักประกันสุขภาพถ้วนหน้า และสวัสดิการรักษายาบาลข้าราชการและครอบครัว) รวมทุ				
2	หน่วย: รายโรค				
3	กลุ่มสาเหตุป่วย 298 โรค	2556			
4		รวม	ชาย	หญิง	
5	รวมทุกกลุ่มโรค	17,999,153	8,338,795	9,660,358	16,4
6	อหิวาตกโรค	227	109	118	
7	ใช้รากสาดน้อยและใช้รากสาดเทียม	5,869	3,090	2,779	
8	โรคบิดจากเชื้อซีเกลลา	365	162	203	
9	โรคบิดมีตัว	3,715	1,771	1,944	
10	อาการท้องร่วง กระเพาะและลำไส้อักเสบ ซึ่งสันนิษฐานว่าเกิดจากการติดเชื้อ	417,287	192,068	225,219	
11	โรคลำไส้อักเสบอื่น ๆ	58,431	28,165	30,266	
12	วัณโรคทางเดินหายใจ	69,033	47,387	21,646	
13	วัณโรคอื่น ๆ	18,056	11,116	6,940	
14	กาฬโรค	0	0	0	
15	โรคติดเชื้อบรูเซลลา	10	10	0	
16	โรคเรื้อน	455	298	157	
17	บาดทะยักในเด็กแรกเกิด	3	1	2	
18	บาดทะยักอื่น ๆ	268	167	101	

Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector_05_15_TH_.xlsx

From: <http://statbbi.nso.go.th/staticreport/page/sector/th/05.aspx>

แหล่งข้อมูลสาธารณะ (ในประเทศไทย)

- สำนักงานสถิติแห่งชาติ
- Government Open Data (<https://data.go.th>)
- เฉพาะด้าน เช่น TPMAP (<https://www.tpmap.in.th>)
- หน่วยงานต่างมักมีหลายระบบ ต้องลองค้นหา
 - กระทรวงพาณิชย์ (<https://data.moc.go.th/> มี api <https://data.moc.go.th/developer>)
 - กระทรวงศึกษาธิการ (<https://ods.moe.go.th:8888/>) + API?
 - ศูนย์ข้อมูลอุบัติเหตุ (<http://www.thairsc.com/>)
 - กรมประมง (หนังสือสถิติการประมง <https://www.fisheries.go.th/strategy-stat/index.php/document-public>)
 - กระทรวงสาธารณสุข (สถิติสาธารณสุข / มีหลายระบบ)
 - กระทรวงคมนาคม (<http://www.motoc.mot.go.th/stat/>)
 - กระทรวงการท่องเที่ยว (https://www.mots.go.th/more_news_new.php?cid=411)
 - สถิติข้อมูลเกี่ยวกับการจัดการน้ำ

แหล่งข้อมูล (นานาชาติ)

- WHO (
<https://www.who.int/healthinfo/statistics/en/>)
- World bank (<https://data.worldbank.org/>)
- PISA (<https://www.oecd.org/pisa/data/>) มี data explorer

ข้อมูลจาก API

- การอ่านข้อมูล
 - curl
 - เขียนโปรแกรม (ด้วยภาษาเช่น python, ruby, etc..)

กิจกรรม 3

- แบ่งกลุ่มใหม่ตามความสนใจ
 - เลือกแหล่งข้อมูล 2 – 3 ที่
 - สืบหาข้อมูล
 - จากแหล่งข้อมูลดังกล่าว
 - สามารถนำมาศึกษาความสัมพันธ์ของอะไรได้บ้าง?
 - อาจจะนำมาตอบคำถามอะไรได้บ้าง?