

Introduction to Data Science

โรงเรียนมหิดลวิทยานุสรณ์

วิทยาศาสตร์ข้อมูลสำหรับการประยุกต์ใช้ในชีวิตจริง (Data Science for Real World Applications)

คำอธิบายรายวิชา

ศึกษาการทำข้อมูลให้เป็นภาพ (data visualization) การเขียนโปรแกรมเพื่อการวิเคราะห์ข้อมูลเบื้องต้น วิทยาการหรือองค์ความรู้ในปัจจุบันทางด้านวิทยาการคอมพิวเตอร์ที่เกี่ยวข้องกับการประมวลผลและวิเคราะห์ข้อมูล การวิเคราะห์ปัญหาที่เกี่ยวข้อง การนำความรู้และแนวคิดเชิงคำนวณไปประยุกต์ใช้เพื่อแก้ปัญหาและพัฒนาชิ้นงานที่บูรณาการกับวิชาอื่น

ผลการเรียนรู้

1. เลือกใช้เครื่องมือและทำข้อมูลให้เป็นภาพ (data visualization) ได้
2. เขียนโปรแกรมเพื่อการประมวลผลและวิเคราะห์ข้อมูลเบื้องต้นได้
3. อธิบายหลักการและทฤษฎีที่เกี่ยวข้องกับองค์ความรู้หรือวิทยาการคอมพิวเตอร์ในปัจจุบันได้
4. วิเคราะห์ปัญหาโดยใช้หลักการและทฤษฎีที่เกี่ยวข้องกับองค์ความรู้หรือวิทยาการคอมพิวเตอร์ในปัจจุบันได้
5. ประยุกต์ใช้หลักการหรือทฤษฎีที่เกี่ยวข้องกับองค์ความรู้หรือวิทยาการคอมพิวเตอร์ในปัจจุบันเพื่อแก้ปัญหา และพัฒนาชิ้นงานที่มีการบูรณาการกับวิชาอื่นอย่างสร้างสรรค์และเชื่อมโยงกับชีวิตจริง

เวลาที่ใช้ในการเรียน

รูปแบบการเรียน	คำอธิบาย	จำนวนเวลา
Synchronous	ช่วงเวลาที่นักเรียนทุกคนเข้าเรียนกับอาจารย์พร้อมกัน	~ 70%
Asynchronous	ช่วงเวลาที่นักเรียนเข้าเรียนแต่ไม่จำเป็นต้องพร้อมกัน	~ 30%
การบ้าน	ทำงานที่ได้รับมอบหมาย	ไม่เกิน 45 นาที / สัปดาห์
คาบพิเศษ	วิชานี้มีจำนวน 1.5 หน่วยกิต จึงมีสัปดาห์ที่มี 4 คาบ สลับกับ 2 คาบ จะเรียก 2 คาบที่เพิ่มขึ้นมาในสัปดาห์ที่มี 4 คาบว่า คาบพิเศษ	

ตารางโดยคร่าว

	สัปดาห์	action	format	concept	คาบพิเศษ	หมายเหตุ
23 พฤษภาคม	1	concept/big picture	lecture	intro to data science+python recap	recitation	
30 พฤษภาคม	2	reading errors	lab	intro to data science+python recap	recitation	TOI
6 มิถุนายน	3	google sheet	lab	available tools	recitation	
13 มิถุนายน	4	numpy	lab	reading documentations	recitation	
20 มิถุนายน	5	pandas	lab	data acquisition	recitation	
27 มิถุนายน	6	numpy+pandas	lab	data cleansing	recitation	

ตารางโดยคร่าว

	สัปดาห์	action	format	concept	คาบพิเศษ	หมายเหตุ
4 กรกฎาคม	7	plotly.express	lab	data visualization	recitation	
11 กรกฎาคม	8	example	meeting	project working time	project working time	ออนไลน์
18 กรกฎาคม	9	example	meeting	project working time	project working time	ออนไลน์
25 กรกฎาคม	10		break			TChO/break
1 สิงหาคม	11	example	meeting	project working time	project working time	ออนไลน์
8 สิงหาคม	12	pandas+altair	lab	data science workflow	recitation	

ตารางโดยคร่าว

	สัปดาห์	action	format	concept	คาบพิเศษ	หมายเหตุ
15 สิงหาคม	13	case study	lab	ethics	project working time	
22 สิงหาคม	14	exam	lecture	exam	project working time	
29 สิงหาคม	15	sklearn	lab	machine learning concept	project working time	
5 กันยายน	16	example	meeting	project working time	project working time	
12 กันยายน	17		presentation	project presentation	project working time	
19 กันยายน	18		presentation	project presentation		

รูปแบบการเรียนในห้อง

Format	คำอธิบาย
Lecture	บรรยายเนื้อหา
Lab	บทปฏิบัติในห้องเรียน ซึ่งจะมีการ Checkoff เก็บ คะแนนกับครูผู้สอนในคาบ
Meeting	นักเรียนนัดพบกับครูผู้สอนเป็นกลุ่ม
Presentation	นักเรียนนำเสนองาน
Recitation	ทบทวนเนื้อหา หรือบทปฏิบัติการเดิมที่ผ่านมา

การประเมินผล

คะแนน	ส่วน	ประกอบด้วย	ส่วนย่อย
70	คะแนนเก็บ	40	lab (checkoff)
		30	project
20	สอบ		
10	จิตพิสัย		

แผ่นย่อ



1

แผนการจัดการเรียนรู้และแผนการประเมินผลการเรียนรู้ ฉบับย่อ

รายวิชา ว30183 วิทยาศาสตร์ข้อมูลสำหรับการประยุกต์ใช้ในชีวิตจริง (Data Science for Real World Applications)

จำนวน 1.5 หน่วยกิต เวลาเรียน 3 คาบ/สัปดาห์ ชั้นมัธยมศึกษาปีที่ 5

ภาคเรียนที่ 1 ปีการศึกษา 2565

อาจารย์ผู้สอน 1. อ.ทศพร แสงจำ (ผู้ประสานงาน) 2. อ.บุญนที ศักดิ์บุญญารัตน์ 3. อ.เลขาขวัญ งามประสิทธิ์
4. อ.สิทธิโชค โสมอ่ำ 5. อ.ศิริประภา เรวก 6. อ.อมรศรี อมรวิชรพงศ์



1. คำอธิบายรายวิชา

ศึกษาการทำข้อมูลให้เป็นภาพ (data visualization) การเขียนโปรแกรมเพื่อการวิเคราะห์ข้อมูลเบื้องต้น วิทยาการหรือองค์ความรู้ในปัจจุบันทางด้านวิทยาการคอมพิวเตอร์ ที่เกี่ยวข้องกับการประมวลผลและวิเคราะห์ข้อมูล การวิเคราะห์ปัญหาที่เกี่ยวข้อง การนำความรู้และแนวคิดเชิงคำนวณไปประยุกต์ใช้เพื่อแก้ปัญหาและพัฒนาชิ้นงานที่บูรณาการกับวิชาอื่น

2. ผลการเรียนรู้

1. เลือกใช้เครื่องมือและทำข้อมูลให้เป็นภาพ (data visualization) ได้
2. เขียนโปรแกรมเพื่อการประมวลผลและวิเคราะห์ข้อมูลเบื้องต้นได้
3. อธิบายหลักการและทฤษฎีที่เกี่ยวข้องกับองค์ความรู้หรือวิทยาการคอมพิวเตอร์ในปัจจุบันได้
4. วิเคราะห์ปัญหาโดยใช้หลักการและทฤษฎีที่เกี่ยวข้องกับองค์ความรู้หรือวิทยาการคอมพิวเตอร์ในปัจจุบันได้
5. ประยุกต์ใช้หลักการหรือทฤษฎีที่เกี่ยวข้องกับองค์ความรู้หรือวิทยาการคอมพิวเตอร์ในปัจจุบันเพื่อแก้ปัญหาและพัฒนาชิ้นงานที่มีการบูรณาการกับวิชาอื่นอย่างสร้างสรรค์และเชื่อมโยงกับชีวิตจริง

Data Science คืออะไร?

<https://www.facebook.com/bigdata.go.th/posts/332450545140933>

Data Science คืออะไร?

รู้จักคำว่า data scientist และ data science กันตั้งแต่เมื่อไหร่ และรู้จักได้อย่างไร

- รู้จักตอนที่เรียน ป.ตรี จากการที่ต้องวิเคราะห์ข้อมูลในการทำงานวิจัย
- เมื่อประมาณ 7-8 ปีก่อน ยังไม่มีชื่อเรียกผู้ที่ทำงานนี้
- เลยได้เข้ามาทำงานในวงการนี้โดยไม่รู้ตัว



Data Science คืออะไร?

data science คืออะไร

- เป็นวิทยาศาสตร์ที่เกี่ยวข้องกับข้อมูล
- ตั้งสมมติฐาน ทำการทดลอง และสรุปผลโดยการใช้กระบวนการวิเคราะห์ข้อมูล



Data Science คืออะไร?

เป็น data scientist นี่
วัน ๆ ทำอะไรกันบ้าง

- Data scientist ไม่ได้ทำการสร้างโมเดล
ทดสอบโมเดลเท่านั้น
- ต้องออกแบบกระบวนการทั้งหมดตั้งแต่การนำ
ข้อมูลเข้า การเลือกใช้ข้อมูล การเลือกโมเดล
ไปจนถึงการส่งมอบโมเดลเพื่อนำไปใช้งาน



Data Science คืออะไร?

ใช้เครื่องมืออะไรในการวิเคราะห์ข้อมูล

- ใช้เครื่องมือคล้าย ๆ กับ โปรแกรมเมอร์ แต่ที่นิยมคือ Python, pandas, Spark, SQL, Excel และเครื่องมือสำหรับ Machine Learning เช่น TensorFlow, Keras, PyTorch
- เครื่องมือที่สำคัญคือทักษะการเรียนรู้สิ่งใหม่ ๆ
- อย่างกลัวที่จะต้องเรียนรู้สิ่งใหม่ ๆ



Data Science คืออะไร?

เป็น data scientist ใช้สถิติมากแค่ไหน

- เบื้องต้นใช้แค่สถิติพื้นฐานก็เพียงพอ
- ถ้าเข้าใจเรื่องการทดสอบสมมติฐานหรือการแจกแจงข้อมูลจะดีมาก จะช่วยให้เข้าใจข้อมูลและช่วยในการตัดสินใจเลือกใช้โมเดลได้ดียิ่งขึ้น
- ถ้าเข้าใจพวกแคลคูลัสด้วยก็จะทำให้เข้าใจการทำงานของโมเดล



Data Science คืออะไร?

ที่เคยคิดไว้ในการเป็น data scientist กับ
ประสบการณ์จริงที่ได้เจอมา
แตกต่างกันหรือไม่ อย่างไร

- ต้องใช้เวลาค่อนข้างมากในการทำความเข้าใจความต้องการของลูกค้า
- เมื่อความต้องการชัดเจนแล้วถึงค่อยลงมือทำโมเดล
- แต่ไม่ใช่ที่ทำเสร็จแล้วจะเสร็จเลย ต้องทำและปรับปรุงไปเรื่อย ๆ เป็น iterative process



Data Science คืออะไร?

มีสิ่งที่ยากให้ domain expert ทำความ
เข้าใจใน data scientist บ้างมัย

- อยากให้เข้าใจว่าการทำงานต้องใช้เวลาในการพัฒนา
- ไม่อยากให้พุ่งเป้าไปที่ผลเลย โดยที่ยังไม่เข้าใจปัญหาที่แท้จริง
- Data science ไม่สามารถแก้ไขได้ทุกปัญหา



Data Science คืออะไร?

อยากเป็น data scientist ต้องเตรียมตัว
ยังไงบ้าง

- การเรียนคอร์สออนไลน์เป็นจุดเริ่มต้นที่ดี
- ศึกษาสิ่งที่คนอื่นวิเคราะห์ใน Kaggle
- อ่านโพสต์ใน medium หรือ towards data science
- นำสิ่งที่ได้มาเขียนเป็น โพสต์ตัวเองเพื่อเรียบเรียงความคิดอีกที
- ลองแข่ง hackathon



Data Science In 5 Minutes





information
is beautiful

Information is Beautiful

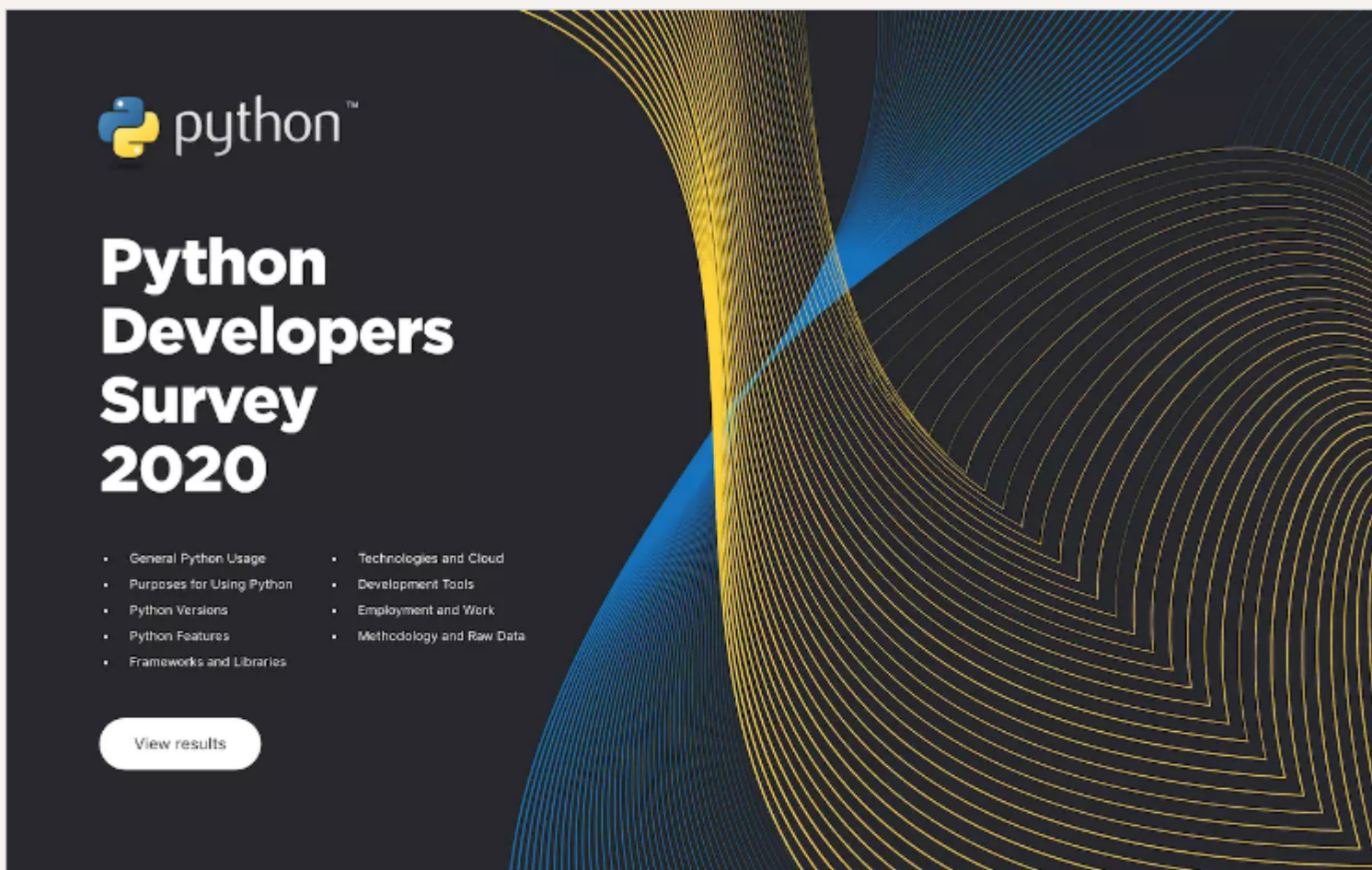
Distilling the world's data, information & knowledge into beautiful infographics & visualizations

 Information is Beautiful

เป้าหมายสำหรับวิชานี้

ต้องการให้นักเรียนสามารถสร้างและพัฒนาชิ้นงานจากข้อมูลในชีวิตจริง

ตัวอย่างรายงานที่มีข้อมูลมาประกอบเพื่อรองรับสมมุติฐาน เช่น

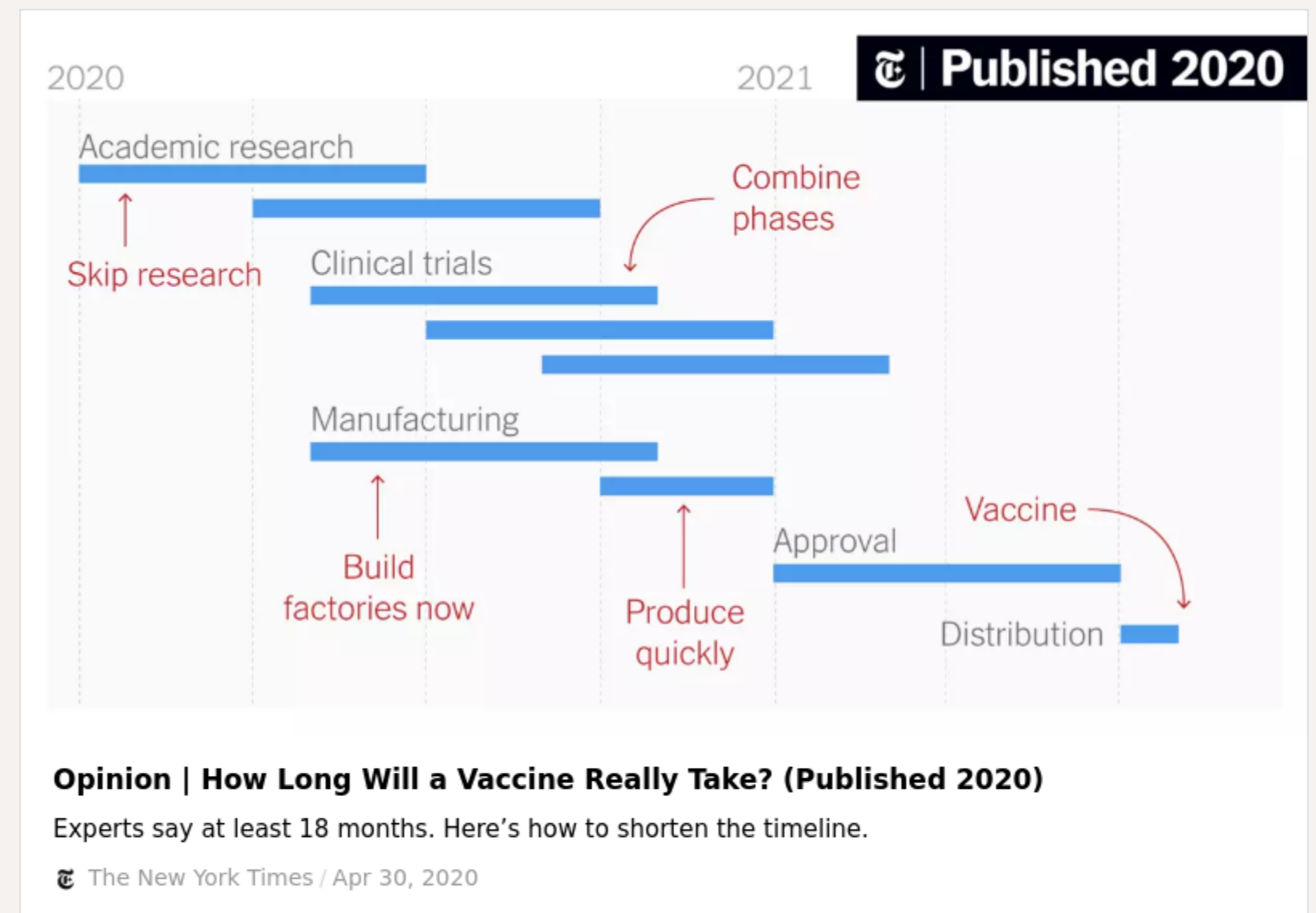


The banner features the Python logo and the text "Python Developers Survey 2020" in a large, bold font. Below the title, there is a list of survey topics: General Python Usage, Technologies and Cloud, Purposes for Using Python, Development Tools, Python Versions, Employment and Work, Python Features, Methodology and Raw Data, Frameworks and Libraries. A "View results" button is located at the bottom left. The background is dark with abstract yellow and blue wavy lines.

Python Developers Survey 2020 Results

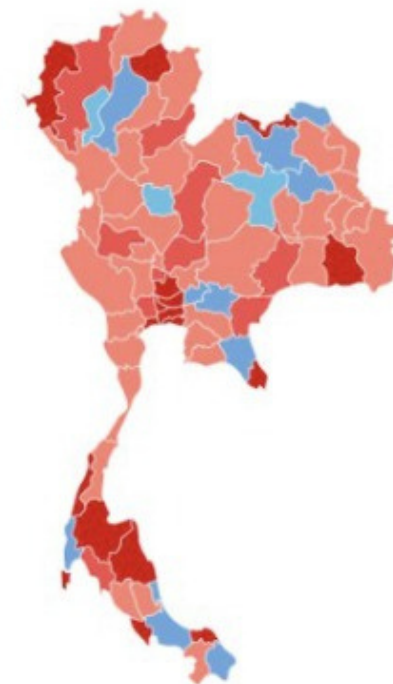
Official Python Developers Survey 2020 Results by Python Software Foundation and JetBrains: more than 28k responses from more than 150 countries.

 JetBrains

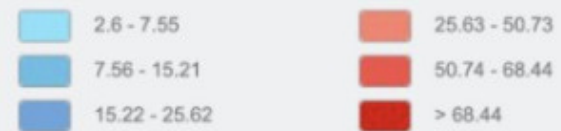


ตัวอย่างผลงานรุ่นพี่

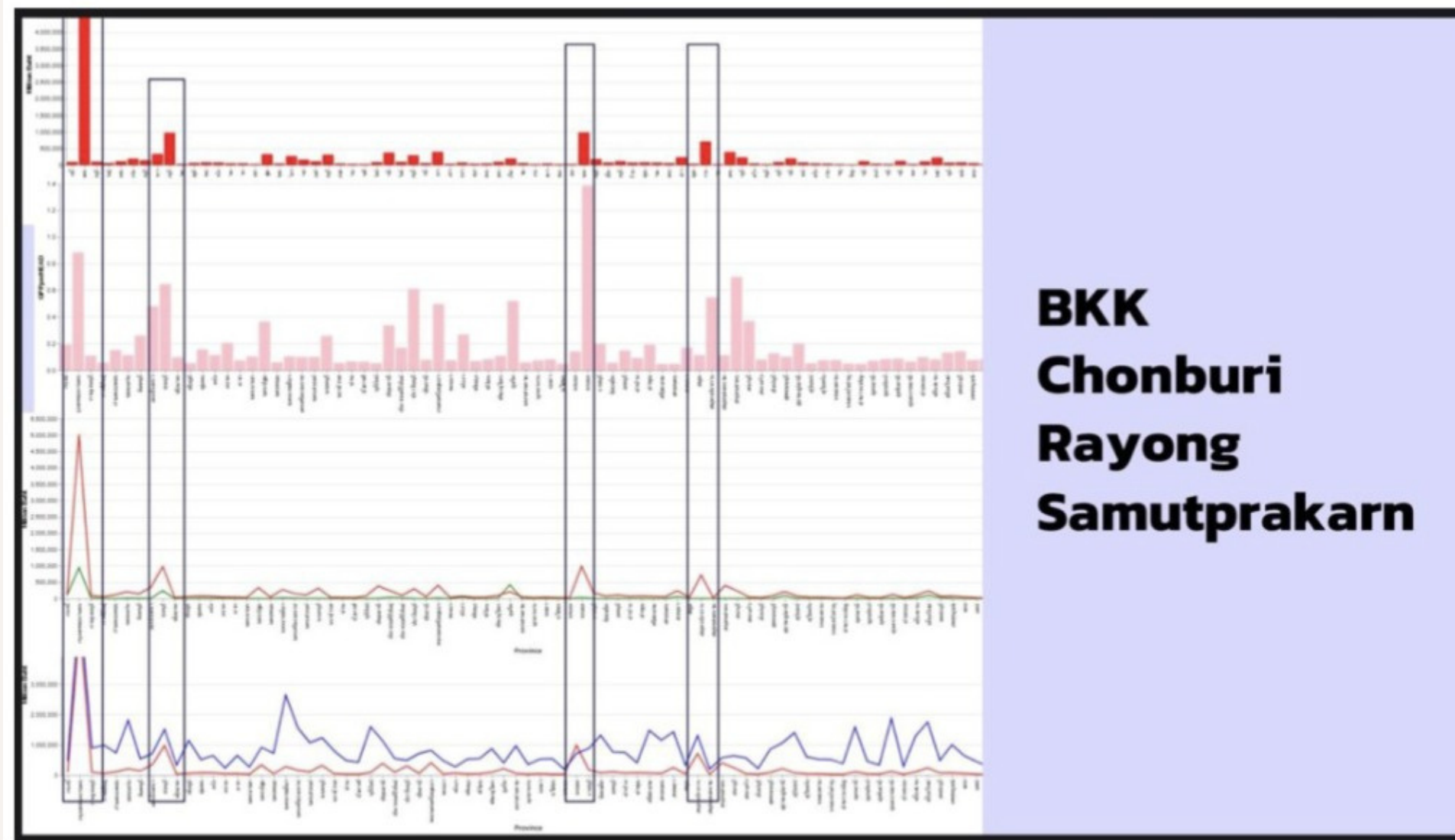
แผนที่ประเทศไทยแสดงจำนวนอุบัติเหตุสะสมของรถยนต์ ต่อจำนวนรถยนต์จดทะเบียน 10,000 คัน แยกราย
จังหวัดตั้งแต่ปีพ.ศ. 2548 ถึงปีพ.ศ. 2558



อัตราการเกิดอุบัติเหตุของรถยนต์ ต่อ จำนวน
รถยนต์จดทะเบียน 10,000 คัน แยกรายจังหวัด



ตัวอย่างผลงานรุ่นพี่

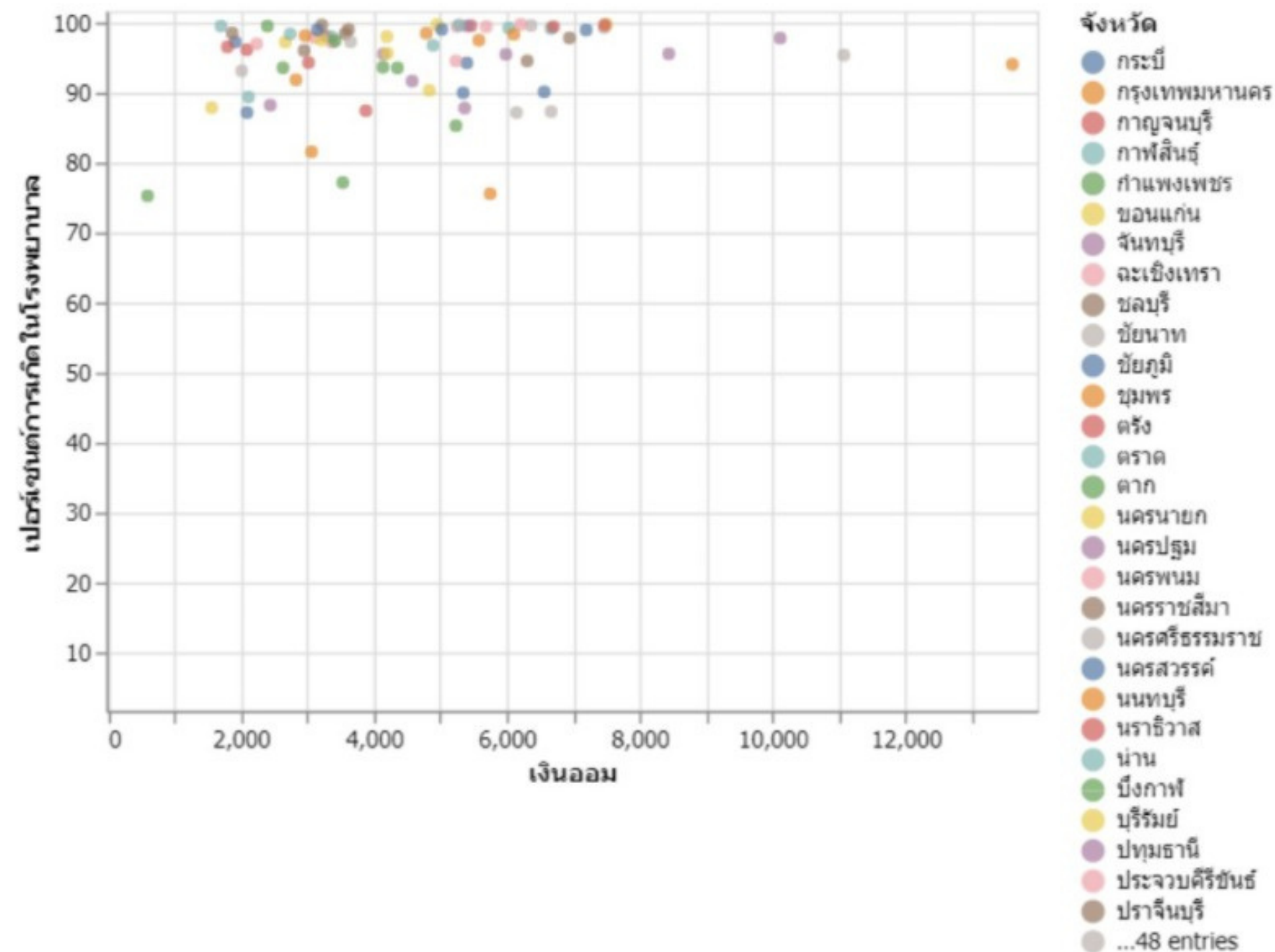


รูปที่ 1 แสดงการเปรียบเทียบระหว่าง กราฟ 4 กราฟ คือ Gross Provincial Product, Gross Provincial Product ต่อคน, Gross Provincial Product กับจำนวนโรงงาน, Gross Provincial Product กับรายได้จากการท่องเที่ยว

ตัวอย่างผลงานรุ่นพี่

สภาพเงินออมบ่งบอกความสามารถในการทำคลอดบุตรได้หรือไม่?

กราฟแสดง10อันดับจังหวัดที่มีเงินออมมากที่สุดช่วงปี 2552-2560



ตัวอย่างผลงานรุ่นพี่

กราฟที่ 2 : แผนภาพแสดงจำนวนนักเรียนทุนในประเทศต่าง ๆ

