การนำเสนอข้อมูลเชิงภาพด้วยจินตทัศน์ข้อมูล

A Visual Data Presentation with Data Visualization

ทัศนันท์ ชูโตศรี*

Thatsanan Chutosri*

สาขาวิชาการจัดการนวัตกรรมดิจิทัลและคอนเทนท์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสวนสุนันทา

Digital Innovation and Content Management Program, Faculty of Science and Technology,

Suan Sunandha Rajabhat University

e-mail: Thatsanan.ch@ssru.ac.th

Received: March 24, 2021; Revised: May 12, 2021; Accepted: May 13, 2021

บทคัดย่อ

จินตทัศน์ข้อมูล เป็นการนำเสนอข้อมูลเชิงภาพ ช่วยให้สามารถแสดงข้อมูลในปริมาณมากและมีความซับซ้อนให้เข้าใจง่ายขึ้น ช่วยในการคาดคะเนสิ่งที่อาจจะเกิดขึ้นในอนาคต และช่วยในการรับรู้ได้ดียิ่งขึ้น การจินตทัศน์ข้อมูลส่วนใหญ่จะถูกออกแบบมาในรูปแบบแผนภูมิแท่ง แผนภูมิเส้น แผนผัง แผนภูมิการกระจาย กราฟเชิงตอบสนอง แผนที่ เครือข่าย และในรูปแบบอื่น ๆ ประโยชน์ของการจินตทัศน์ข้อมูลทำให้ข้อมูลมีความน่าสนใจขึ้น เข้าใจง่าย เห็นภาพได้ชัดเจน ง่ายต่อการจดจำ และสร้างมูลค่าให้กับข้อมูล ซึ่งเป็นสิ่งที่มีความสำคัญมากในยุคดิจิทัลที่มีข้อมูลมหาศาล บทความวิชาการนี้นำเสนอเกี่ยวกับความหมายของจินตทัศน์ข้อมูล รูปแบบความสัมพันธ์ของข้อมูล ประเภทของแผนภูมิที่ใช้ในการจินตทัศน์ข้อมูล และกระบวนการจินตทัศน์ข้อมูล เพื่อสร้างความเข้าใจเกี่ยวกับการจินตทัศน์ข้อมูลได้ดียิ่งขึ้น

คำสำคัญ: ข้อมูล จินตทัศน์ จินตทัศน์ข้อมูล การนำเสนอข้อมูล ข้อมูลเชิงภาพ แผนภูมิ กราฟ

Abstract

Data visualization is a visual data presentation with image. This makes it easier to display large volumes and complex data. It helps to predict what might happen in the future and enhances perceptions efficiency. Most data visualizations are designed in the form of a bar chart, line chart, flow chart, distribution chart scatterplots, surface plots, maps, networks, and other forms. The benefits of visualizing data are making information more interesting, easy to understand, and clearly visible. It also makes it easy to remember and creates value for information which is extremely crucial in an enormous digital age. This academic article presents the definition of the meaning of imaginary data, data correlation model, types of charts used to data visualization and the process of data visualization for better understanding of data visualization.

Keywords: Data, Visualization, Data Visualization, Data Presentation, Visual Information, Charts, Graphs

ความนำ

กราฟิกเชิงสถิติและจินตทัศน์ข้อมูล เป็นพัฒนาการด้านสถิติในรูปแบบโมเดลความสัมพันธ์ (Friendly, 2008) ซึ่งส่วนใหญ่จะเข้าใจว่ากราฟิกเชิงสถิติและการจินตทัศน์ข้อมูลเป็นพัฒนาการที่ทันสมัยในเชิงสถิติ แต่ในความเป็นจริงแล้วการนำเสนอข้อมูลด้วยภาพมีรากเหง้ามาตั้งแต่ในอดีต ในประวัติศาสตร์มีการสร้างแผนที่เป็นการบรรยายด้วยภาพ ต่อมา

ได้ทำแผนที่ สถิติ และกราฟิกเชิงสถิติที่ใช้มานานแล้ว โดยเฉพาะในสาขาการแพทย์ สาขาวิทยาศาสตร์ ต่อมานำมาใช้ในหลากหลายแขนง และนำมาใช้ในการคิดเชิงสถิติและการรวบรวมข้อมูลเพื่อการวางแผนทางการค้าจนถึงศตวรรษที่ 19 และมีการพัฒนาส่งผลให้การจินตทัศน์ข้อมูลมีการนำมาใช้กันอย่างกว้างขวางในปัจจุบัน (Friendly, 2008, p. 1)

ในปัจจุบันการจินตทัศน์ข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็วและซับซ้อนมากขึ้นในทุก ๆ วัน เนื่องจากข้อมูลมีสารสนเทศที่เพิ่มขึ้นและมีความซับซ้อนมากขึ้นจากการใช้โซเชียลมีเดียที่เพิ่มขึ้น การใช้โทรศัพท์มือถือที่แพร่หลาย และการให้บริการในรูปแบบดิจิทัล จึงทำให้มีข้อมูลที่เกี่ยวกับกิจกรรมของมนุษย์ในการใช้เทคโนโลยีมากขึ้น ข้อมูลจึงมีคุณค่าอย่างมหาศาลในการวิเคราะห์แนวโน้ม และใช้ข้อมูลขนาดใหญ่ (Big Data) เชื่อมโยงระหว่างเหตุการณ์ต่าง ๆ ดังนั้นการจินตทัศน์ข้อมูลจึงเป็นกลไกที่มีประสิทธิภาพในการนำเสนอข้อมูลที่เข้าใจสารสนเทศได้แบบเรียลไทม์ และทุกบริษัทมีข้อมูลเพื่อสื่อสารกับลูกค้า ผู้บริหาร และเพื่อช่วยจัดการองค์กรไม่ใช่แค่เพียงการวิจัยเท่านั้น การตีความด้วยการจินตทัศน์ข้อมูลสามารถสื่อความหมายและนำมาเปลี่ยนเป็นความรู้ได้ เพราะในยุคของข้อมูลที่ภาพและเนื้อหามีบทบาทสำคัญในทุกช่วงเวลาของชีวิต (Matias, n.d., p. 4) สารสนเทศสามารถนำเสนอได้หลายวิธี ซึ่งแต่ละวิธีสามารถนำเสนอข้อมูลเชิงลึกที่มีความเฉพาะเจาะจง การทำงานด้านข้อมูลสิ่งที่สำคัญคือ ต้องแยกแยะและทำความเข้าใจเรื่องราวที่ต้องการจะบอกให้สัมพันธ์กับสิ่งที่ต้องการนำเสนอ การมีความรู้ด้านจินตทัศน์ข้อมูลจะช่วยแสดงภาพที่เหมาะสมเพื่อสื่อความหมายได้ดี (Hubspot & Visage, n.d., p. 2)

การจินตทัศน์ข้อมูลจึงมีความสำคัญ เนื่องจากสมองของมนุษย์จะประมวลผลข้อมูลด้วยแผนภูมิ (Charts) หรือ กราฟ (Graphs) ในการแสดงข้อมูลที่มีความซับซ้อนและมีจำนวนมากได้ง่ายกว่าข้อมูลที่อยู่ในรูปแบบตารางหรือรายงาน การจินตทัศน์ข้อมูลจึงเป็นวิธีที่ง่ายและรวดเร็วในการนำเสนอข้อมูลและเป็นสากล (“Data Visualization: What,” 2021) และมีผลการศึกษายืนยันโดย SHIFT ในเรื่องการ Disruptive Learning แสดงให้เห็นว่า คนโดยทั่วไปจะประมวลภาพได้เร็วกว่าตารางหรือข้อความได้ถึง 60,000 เท่า และสมองจะจดจำภาพเหล่านั้นได้ดีกว่าในระยะยาว และยังพบว่าหลังจากนั้นสามวันได้วิเคราะห์ข้อมูลที่ผู้เข้าร่วมวิจัยสามารถจดจำได้ ข้อมูลที่เป็นลายลักษณ์อักษรหรือคำพูดจดจำได้ 10% และ 20% เมื่อเทียบข้อมูลจากภาพจดจำได้ 65% และจิตของมนุษย์สามารถมองเห็นภาพได้เพียง 13 มิลลิวินาที ดวงตาของเราสามารถรับข้อความภาพได้ 36,000 ข้อความต่อชั่วโมง 40% ของเส้นประสาท เชื่อมต่อกับเรตินา ทั้งหมดนี้บ่งบอกได้ว่ามนุษย์สามารถประมวลผลข้อมูลภาพดีกว่า และอยู่ในความทรงจำได้ยาวนาน ดังนั้นการรายงานด้วยภาพโดยใช้รูปภาพจึงเป็นวิธีการสื่อสารข้อมูลที่มีประสิทธิภาพมากกว่าข้อความหรือตาราง อีกทั้งยังใช้พื้นฐานน้อยกว่า การสื่อความหมายด้วยภาพจึงทำให้มีความน่าสนใจมากขึ้น ง่ายต่อการนำมาใช้งานและการจดจำ (Matias, n.d., p. 6) การจินตทัศน์ข้อมูลจึงมีความสำคัญในการนำเสนอข้อมูลที่มีปริมาณมากและซับซ้อน เพื่อให้ง่ายต่อการเข้าใจมากยิ่งขึ้น

บทความนี้มีวัตถุประสงค์เพื่อถ่ายทอดองค์ความรู้เกี่ยวกับจินตทัศน์ข้อมูล (Data Visualization) และเพื่อสร้างความเข้าใจเกี่ยวกับการจินตทัศน์ข้อมูลได้ดียิ่งขึ้น บทความนี้มีหัวข้อสำคัญ ได้แก่ ความหมายของจินตทัศน์ข้อมูล รูปแบบความสัมพันธ์ของข้อมูล ประเภทของแผนภูมิ กระบวนการจินตทัศน์ข้อมูล และเทคโนโลยีที่นำมาประยุกต์ใช้ในการจินตทัศน์ข้อมูล

ความหมายของจินตทัศน์ข้อมูล

คำจำกัดความของจินตทัศน์ข้อมูลยังไม่มีการจำกัดความที่ชัดเจน เนื่องจากมีคำในภาษาอังกฤษหลายคำที่มีความหมายที่ใช้ใกล้เคียงกับ Data Visualization เช่นคำว่า Information Visualization และ Infographics ที่ใช้แทนกันได้ แต่มีนักออกแบบ ผู้ปฏิบัติงาน และนักวิชาการได้ถกเถียงกันหลายท่านยืนยันว่าคำเหล่านี้เป็นคำที่พ้องความหมาย และพยายามที่จะแยกแยะความแตกต่างของการแสดงข้อมูลด้วยการเชื่อมโยงกับข้อมูล

(Strecker, 2012, p. 4) จากการศึกษาความหมายของคำว่า จินตทัศน์ข้อมูล (Data Visualization) ได้มีผู้เชี่ยวชาญนิยามความหมาย ไว้ดังต่อไปนี้

Nediger (2020) ได้ให้ความหมายของคำว่า จินตทัศน์ข้อมูล (Data Visualization) คือการนำเสนอข้อมูลหรือสารสนเทศ เป้าหมายของจินตทัศน์ข้อมูลคือการสื่อสารข้อมูลหรือสารสนเทศให้มีความชัดเจนและมีประสิทธิภาพสำหรับผู้อ่าน โดยทั่วไปจะแสดงข้อมูลในรูปแบบแผนภูมิ อินโฟกราฟิก แผนภาพ หรือแผนที่

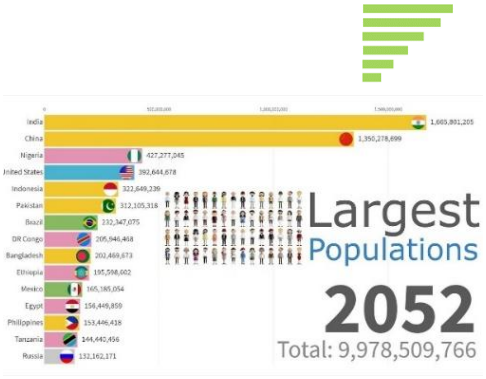
วรวิกา วัฒนสุนทร และชุติมณฑน์ รักนะ (2560, หน้า 87) การสร้างภาพจินตทัศน์ คือ เครื่องมือเทคนิควิธีการ กระบวนการแปลงข้อมูลสารสนเทศและองค์ความรู้ โดยใช้คอมพิวเตอร์ในการนำเสนอผลลัพธ์แบบแผนภาพ แผนภูมิ แผนผัง กราฟิกโครงสร้าง แผนที่ หรือรูปแบบต่าง ๆ เพื่อช่วยในการสนับสนุนการตัดสินใจของผู้ใช้ อีกทั้งยังช่วยสนับสนุนงานวิจัยในวงกว้าง เช่น การวิเคราะห์ข้อมูลสารสนเทศ การสำรวจข้อมูลสารสนเทศ การทำนายแนวโน้มช่วงเวลา การเล่าเรื่องราวที่ผ่านมา เป็นต้น

สรุปได้ว่าการจินตทัศน์ข้อมูล (Data Visualization) หมายถึง การนำเสนอข้อมูลหรือสารสนเทศด้วยภาพหรือภาพกราฟิก โดยใช้สี รูปร่างหรือรูปทรงสัญลักษณ์ และใช้กระบวนการทางเทคนิคในการแปลงข้อมูลหรือสารสนเทศให้กลายเป็นภาพหรือภาพกราฟิก เพื่อช่วยในการมองเห็น ทำให้เข้าใจข้อมูลที่มีความซับซ้อนได้ง่ายขึ้น การนำเสนอส่วนใหญ่จะนำเสนอเป็นแผนภาพ แผนภูมิ แผนผัง ภาพกราฟิก แผนที่ หรืออยู่ในรูปแบบอื่น ๆ ซึ่งนำมาช่วยในการตัดสินใจ เล่าเรื่อง และวิเคราะห์แนวโน้มต่าง ๆ เป็นต้น

รูปแบบความสัมพันธ์ของข้อมูล

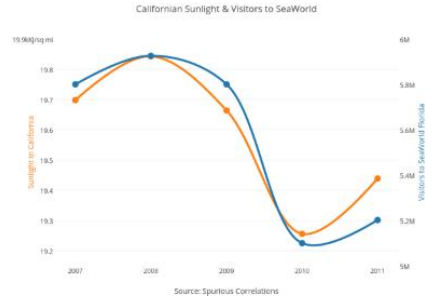
รูปแบบความสัมพันธ์ของข้อมูลในการนำมาจินตทัศน์ มีหลากหลายรูปแบบความสัมพันธ์ ซึ่งจะยกตัวอย่างมานำเสนอบางส่วน (Matias, n.d., p. 10) ดังนี้

1. จัดอันดับ (Ranking) การแสดงภาพข้อมูลที่มีความสัมพันธ์กันตั้งแต่สองค่าขึ้นไป



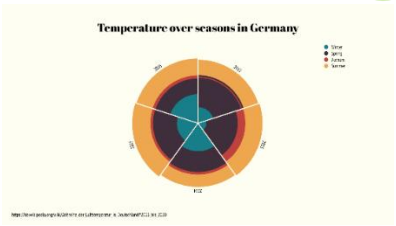
ภาพที่ 1 ตัวอย่างการจัดอันดับ Most Populated Countries in the Future 2020-2100
ที่มา (Rana, 2020)

2. ความสัมพันธ์ (Correlation) ความสัมพันธ์ของข้อมูลที่มีสองตัวแปรขึ้นไปที่สามารถแสดงให้เห็นถึงความสัมพันธ์เชิงบวกหรือเชิงลบ



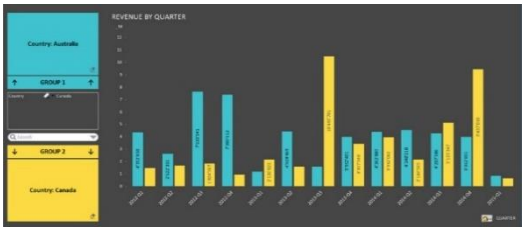
ภาพที่ 2 แสดงความสัมพันธ์ระหว่าง California Sunlight & Visitors to SeaWorld
ที่มา (“Californian sunlight,” 2013)

3. ความสัมพันธ์บางส่วนและทั้งหมด (Partial and Total Relationships) แสดงชุดข้อมูลย่อยเปรียบเทียบกับผลรวมทั้งหมด



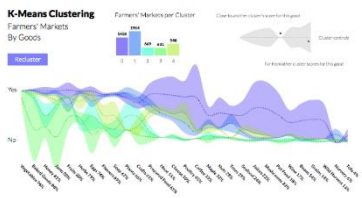
ภาพที่ 3 ตัวอย่างแผนภูมิความสัมพันธ์บางส่วนและทั้งหมด
ที่มา (“Temperature over,” 2021)

4. การเปรียบเทียบ (Comparisons) การแสดงภาพที่เปรียบเทียบข้อมูลเชิงปริมาณ



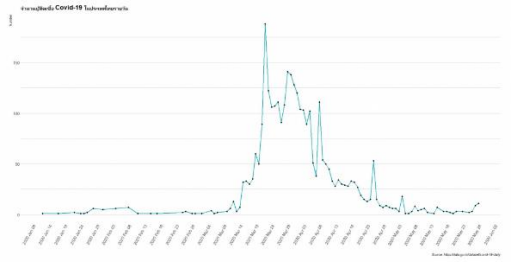
ภาพที่ 4 ตัวอย่างการเปรียบเทียบระหว่าง 2 ประเทศ
ที่มา (“Data visualization best,” 2020)

5. การหักเห (Deviation) ตรวจสอบว่าข้อมูลแต่ละจุดที่มีความเกี่ยวข้องกัน โดยเฉพาะอย่างยิ่งจุดที่ชี้ให้เห็นมูลค่าที่แตกต่างจากค่าเฉลี่ย



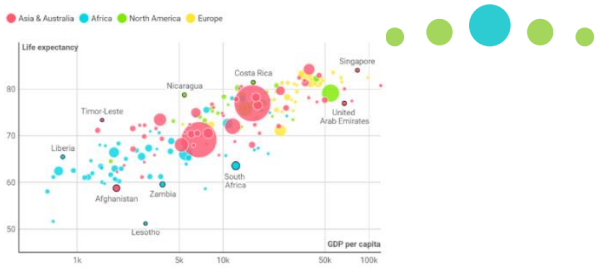
ภาพที่ 5 K-Means Centroid Deviation
ที่มา (Susie, 2020)

6. จัดลำดับตามเวลา (Series Time) แสดงการเปลี่ยนแปลงของข้อมูลในแต่ละช่วงเวลา



ภาพที่ 6 กราฟแสดงจำนวนผู้ติดเชื้อ COVID-19
ที่มา (ยุรนันท์ จามจุรี, 2563)

7. การกระจาย (Distribution) การแสดงภาพการกระจายของข้อมูลเชิงพื้นที่

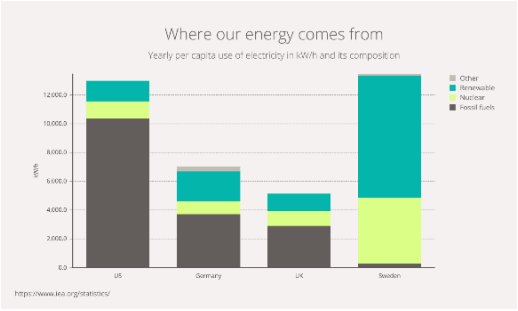


ภาพที่ 7 ตัวอย่างการแสดงภาพการกระจายของข้อมูลเชิงพื้นที่
ที่มา (Philp, 2020)

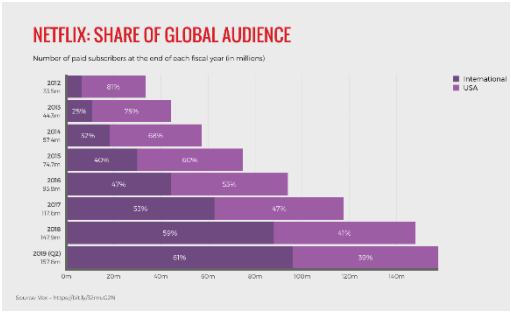
ประเภทของแผนภูมิที่ใช้ในการจินตทัศน์ข้อมูล

ประเภทของแผนภูมิที่ใช้ในการนำเสนอจินตทัศน์ข้อมูลที่พบบ่อย มีดังนี้

1. แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้ง ที่นิยมแสดงออกมาในรูปแท่งสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับใช้ในการเปรียบเทียบจำนวนของข้อมูลในแต่ละชุด (“การเลือกรูปแบบ,” 2564)

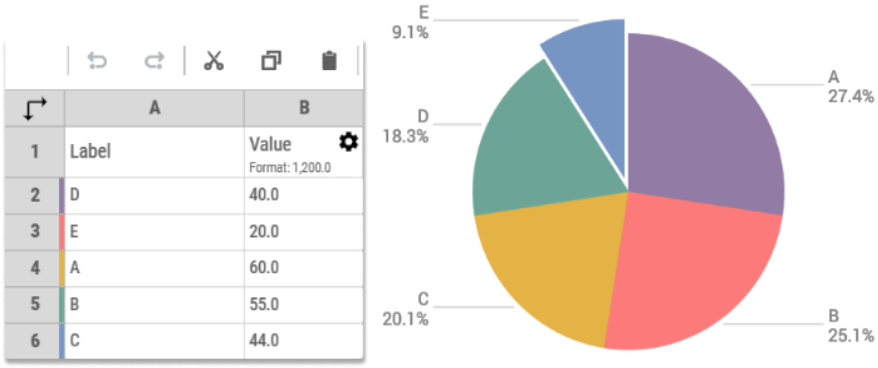


ภาพที่ 8 แผนภูมิแท่งแนวตั้ง (Vertical Charts)
ที่มา (“Where our energy,” 2021)



ภาพที่ 9 แผนภูมิแนวนอน (Horizontal Charts)
ที่มา (“Netflix: Share,” 2021)

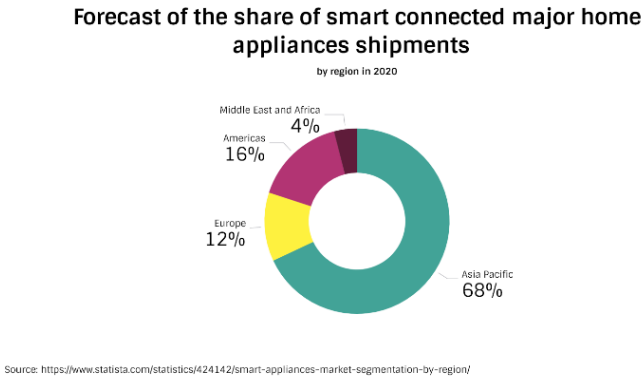
2. แผนภูมิวงกลม (Pie Charts) เหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่ายและสวยงาม แต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยากเพราะต้องใช้หลายสีในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share) ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น (“การเลือกรูปแบบ,” 2564)



	A	B
1	Label	Value Format: 1,200.0
2	D	40.0
3	E	20.0
4	A	60.0
5	B	55.0
6	C	44.0

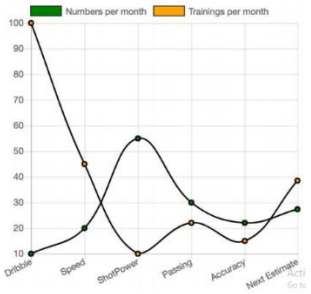
ภาพที่ 10 แผนภูมิวงกลม (Pie Charts)
ที่มา (“Add your data,” 2021)

3. แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลม แต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลาย ๆ ชั้น (“การเลือกรูปแบบ,” 2564)



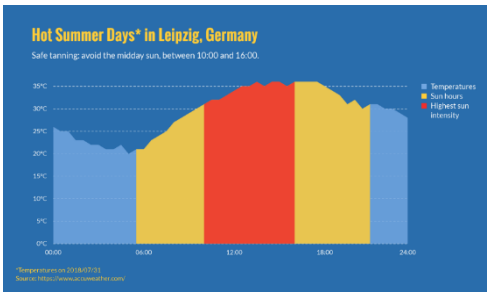
ภาพที่ 11 แผนภูมิโดนัท (Doughnut Charts)
ที่มา (“Forecast of the share,” 2020)

4. แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและแกนนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมิ แผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลา เพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้ (“การเลือกรูปแบบ,” 2564)



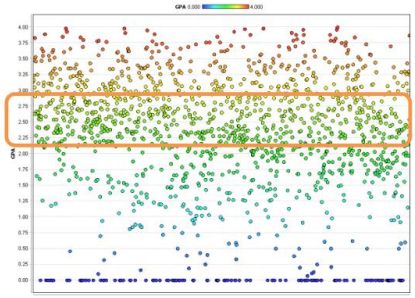
ภาพที่ 12 แผนภูมิเส้น (Line Charts)
ที่มา (Roshdy, Sharaf, Saad, & Abdennadher, 2018)

5. แผนภูมิพื้นที่ (Area Charts) แสดงความสัมพันธ์ด้านเวลา หน้าตาคล้ายแผนภูมิเส้น แต่แตกต่างจากแผนภูมิเส้นที่สามารถแสดงปริมาณ มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล (“การเลือกรูปแบบ,” 2564)



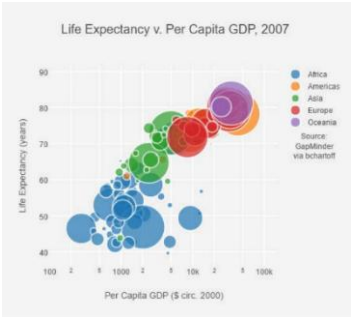
ภาพที่ 13 แผนภูมิพื้นที่ (Area Charts)
ที่มา (“Hot summer,” 2020)

6. แผนภูมิการกระจาย (Scatter Plot Charts) ใช้แสดงความสัมพันธ์ระหว่างตัวแปรข้อมูลอย่างน้อยสองชุด เพื่อใช้แสดงความสัมพันธ์ของข้อมูลที่มีจำนวนมากได้ดีที่สุด



ภาพที่ 14 ตัวอย่างแผนภูมิการกระจาย (Scatter Plot Charts) ของเกรดเฉลี่ย (GPA)
ที่มา (Chertchom, 2017)

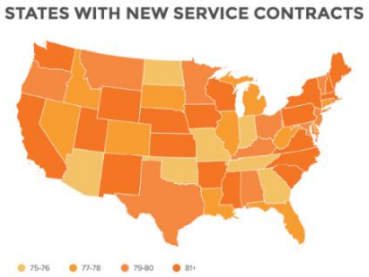
7. แผนภูมิฟอง (Bubble Charts) แผนภูมิฟองเหมาะสำหรับการแสดงข้อมูลขนาดเล็กในรูปแบบของการเปรียบเทียบหรือจัดอันดับความสัมพันธ์ (“การเลือกรูปแบบ,” 2564)



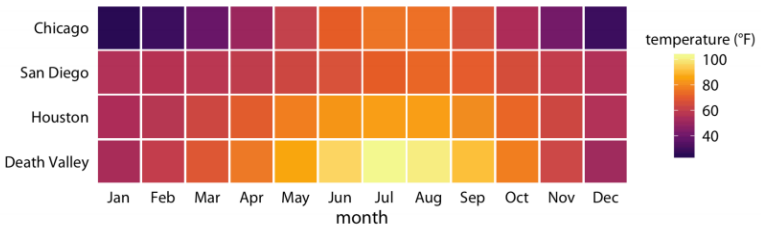
ภาพที่ 15 แผนภูมิฟองกระจาย (Bubble Plot Charts)
ที่มา (Ali, Gupta, Nayak, & Lenka, 2016, p. 658)

ภาพที่ 16 แผนภูมิฟองแผนที่ (Bubble Map Charts)
ที่มา (Ali et al., 2016, p. 658)

8. แผนภูมิความร้อน (Heat Map Charts) แผนภูมิความร้อนแสดงข้อมูลที่ใช้ความเข้มของสี เพื่อแสดงข้อมูลพื้นที่ทางภูมิศาสตร์หรือตารางข้อมูล (“การเลือกรูปแบบ,” 2564)

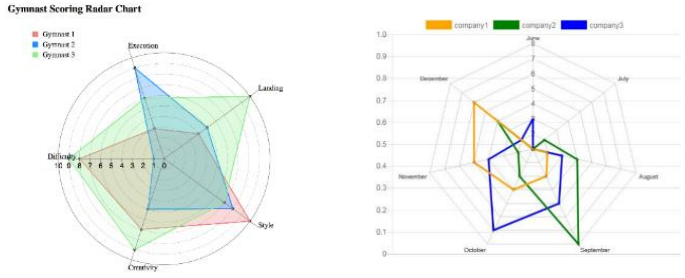


ภาพที่ 17 แผนภูมิความร้อน (Heat Map Charts)
ที่มา (Hubspot & Visage, n.d.)



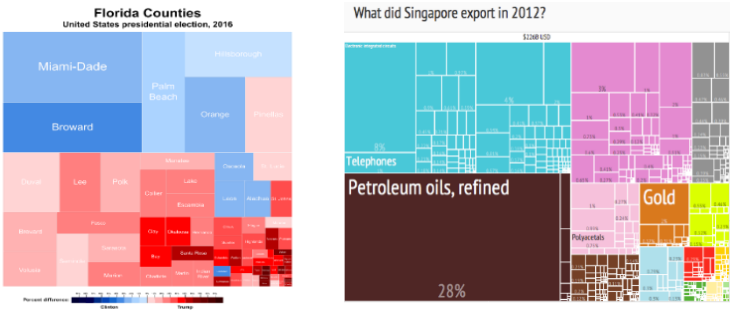
ภาพที่ 18 แผนภูมิความร้อน (Heat Map Charts)
ที่มา (Wilke, 2019)

9. แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูล แต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูล (“การเลือกรูปแบบ,” 2564)



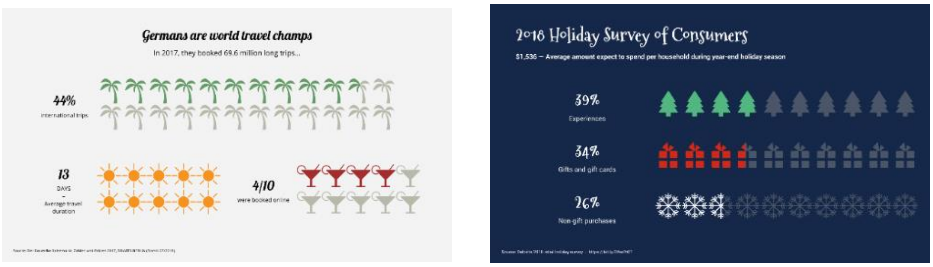
ภาพที่ 19 แผนภูมิเรดาร์ (Radar Charts)
ที่มา (Nowicki & Merenstein, 2016; Roshdy et al., 2018)

10. แผนภูมิต้นไม้ (Tree Maps Charts) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้ (“การเลือกรูปแบบ,” 2564)



ภาพที่ 20 แผนภูมิต้นไม้ (Tree Maps Charts)
ที่มา (Zifan, 2017; Gordon.silvermanaz, 2014)

11. แผนภูมิรูปภาพ (Picture Graph Charts) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และแกนตั้ง แต่เลือกใช้รูปภาพหรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ (“การเลือกรูปแบบ,” 2564)



ภาพที่ 21 แผนภูมิรูปภาพ (Picture Graph Charts)
ที่มา (“Germans are world,” 2020; “2018 Holiday survey,” 2020)

กระบวนการจินตทัศน์ข้อมูล

กระบวนการจินตทัศน์ข้อมูลมีเป้าหมายเพื่อลดความซับซ้อนและแสดงความสัมพันธ์ของข้อมูล มีนักวิชาการหลายท่านได้ทำการศึกษากระบวนการจินตทัศน์ข้อมูล ดังต่อไปนี้

กระบวนการจินตทัศน์ข้อมูล มีหลายขั้นตอนที่เกี่ยวข้องกับการแสดงข้อมูล ตั้งแต่การศึกษาชุดข้อมูลไปจนถึงการพล็อตกราฟ ซึ่งมีขั้นตอน (Joshi, Chawla, & Shukla, 2020, pp. 328-329) ดังต่อไปนี้

1. การวิเคราะห์ปัญหา (Analysing the Problem)

การวิเคราะห์ปัญหา จากการศึกษาชุดข้อมูล โดยศึกษาจากข้อมูลที่ไม่ใช่ตัวเลข (NaN) ที่อยู่ในชุดข้อมูลที่ได้รวบรวมไว้ (Dataset) ในขั้นตอนนี้สามารถเลือกประเภทของภาพในการแสดงชุดข้อมูลที่รวบรวมไว้เพื่อให้แสดงข้อมูลได้อย่างชัดเจนและเข้าใจปัญหาได้ดีขึ้น

2. การทำความสะอาดข้อมูล (Cleaning the Datasets)

การทำความสะอาดข้อมูล Data Cleansing หรือ Data Cleaning หรือ Data Scrubbing คือ กระบวนการตรวจสอบ การแก้ไข หรือการลบ เพื่อให้รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล เพราะหมายถึงความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่น ๆ เป็นต้น จึงทำให้ผู้เชี่ยวชาญหลาย ๆ คนมองว่าการล้างข้อมูลเป็นสิ่งสำคัญที่สุดในการจัดการกับคุณภาพของข้อมูล (“การทำความสะอาดข้อมูล,” 2563)

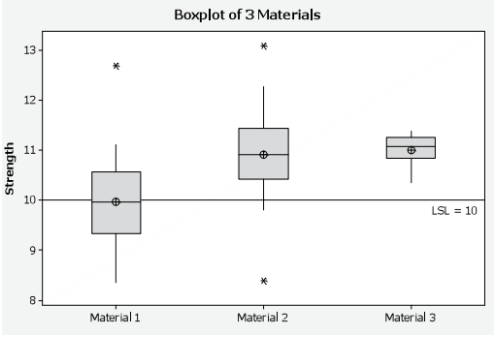
3. การเตรียมข้อมูล (Data Preprocessing)

การเตรียมข้อมูลเป็นขั้นตอนที่สำคัญมากในกระบวนการของการทำงาน ทำหลังจากการทำความสะอาดข้อมูล การเตรียมข้อมูลเพื่อให้ข้อมูลมีความพร้อมก่อนนำไปใช้ ซึ่งข้อมูลที่มีคุณภาพจะต้องมีความถูกต้อง แม่นยำ สม่ำเสมอ และเป็นแบบเดียวกัน เพื่อให้ได้ข้อมูลมีคุณภาพและเป็นมาตรฐาน

4. ขั้นตอนและวิธีการ (Algorithm and Methods)

ขั้นตอนและวิธีการในการแสดงข้อมูล วิธีแสดงข้อมูลออกมาให้อยู่ในรูปแบบจินตทัศน์ข้อมูล (Data Visualization) เป็นวิธีหนึ่งที่มีประสิทธิภาพ ที่สามารถช่วยให้เห็นภาพรวมและการกระจายตัวของข้อมูล และค้นพบนัยบางอย่างของข้อมูลได้ เช่น ค่าความผิดปกติของข้อมูล (Outliers) แนวโน้ม รูปแบบพฤติกรรม หรือความสัมพันธ์อื่น ๆ เป็นต้น ตัวอย่างกราฟที่นิยมใช้ในการแสดงข้อมูล เช่น Boxplot, Line Chart, Scatter และ Histogram เป็นต้น ยกตัวอย่างเช่น

4.1 กราฟรูปกล่อง (Boxplot) เป็นวิธีมาตรฐานสำหรับการทำจินตทัศน์ข้อมูล เป็นกราฟซึ่งแสดงภาพรวมของข้อมูล โดยแสดงให้เห็นทั้งค่ากลางและความผันแปรของข้อมูล นอกจากนี้ Box Plot ยังช่วยในการเปรียบเทียบข้อมูลได้เป็นอย่างดี (จรัล ทรัพย์เสรี, 2553, หน้า 37) สามารถแสดงลักษณะการกระจายตัวของข้อมูลว่ามีลักษณะอย่างไร รวมถึงใช้แสดงค่าความผิดปกติของข้อมูล (Outlier) ได้เป็นอย่างดี

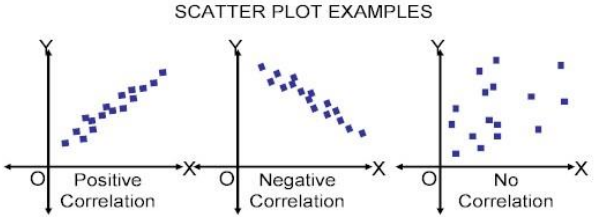


ภาพที่ 22 ตัวอย่างกราฟรูปกล่อง (Boxplot)
ที่มา (จรัล ทรัพย์เสรี, 2553, หน้า 37)

4.2 แผนภูมิเส้น (Line Chart) เป็นการนำเสนอข้อมูล โดยใช้จุดและส่วนของเส้นตรงที่ลากเชื่อมต่อจุด ซึ่งแต่ละจุดจะบอกจำนวนหรือปริมาณของข้อมูล นิยมใช้การนำเสนอข้อมูลแบบแผนภูมิเส้นกับข้อมูลที่มีการเปลี่ยนแปลงอย่างต่อเนื่อง เพราะลักษณะของแผนภูมิเส้นนั้นจะแสดงการเปลี่ยนแปลงของข้อมูลอย่างชัดเจน การอ่านแผนภูมิเส้นต้องอ่านให้สัมพันธ์กับแกนตั้งและแกนนอนด้วย ซึ่งแผนภูมิเส้น มีลักษณะคล้าย ๆ แผนภูมิแท่ง (Bar Chart) เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมิ

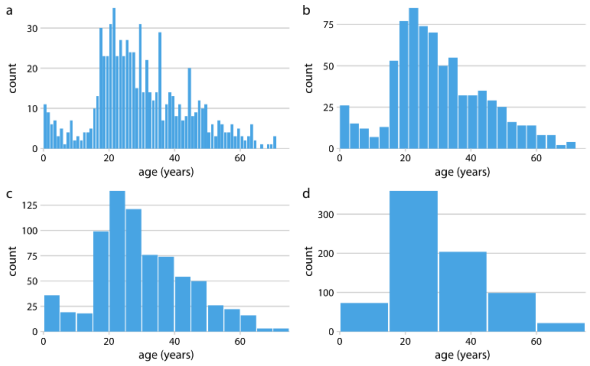
ภาพที่ 23 ตัวอย่างแผนภูมิเส้น (Line Chart)
ที่มา (“การใช้งานแผนภูมิเส้น,” 2563)

4.3 กราฟกระจัดกระจาย (Scatter Plot) เป็นอีกหนึ่งกราฟที่นักวิเคราะห์ข้อมูลนำมาใช้เพื่อแสดงผลข้อมูล โดยการแสดงผลของกราฟจะเป็นการนำเอาจุดตัดกันของค่าในแนวแกน X และแกน Y ของทั้งสองข้อมูลมาแสดง ซึ่งจะช่วยให้เราสามารถเห็นถึงความสัมพันธ์ระหว่างข้อมูลสองตัวแปรว่าสัมพันธ์กันอย่างไร



ภาพที่ 24 ตัวอย่างกราฟกระจัดกระจาย (Scatter Plot)
ที่มา (ณัฐวิภา, 2559)

4.4 ฮีสโตแกรม (Histogram) (Scott, 2008) ฮีสโตแกรม หรือ กราฟแท่ง เป็นหนึ่งในกราฟที่มีความสำคัญในเชิงสถิติ ที่แสดงความสัมพันธ์ระหว่างข้อมูลเป็นหมวดหมู่ที่เรียกว่าชั้นข้อมูลกับความถี่ของข้อมูล เพื่อดูการกระจายของข้อมูล ลักษณะของข้อมูลที่เป็นหมวดหมู่จะเรียงลำดับจากน้อยไปหามากโดยจำนวนหมวดหมู่ของข้อมูลจะจัดตามความเหมาะสม โดยแกนตั้งจะเป็นตัวเลขแสดงความถี่ และแกนนอนจะเป็นข้อมูลคุณสมบัติของสิ่งที่เราสนใจ แท่งกราฟแต่ละแท่งจะมีความกว้างเท่ากันซึ่งเท่ากับความกว้างของชั้นข้อมูล ส่วนความสูงของกราฟแต่ละแท่งนั้นจะสูงเท่ากับจำนวนความถี่ของแต่ละชั้นข้อมูล



ภาพที่ 25 ตัวอย่างฮีสโตแกรม (Histogram)
ที่มา (ณัฐดนัย หวังพระธรรม, 2562)

กระบวนการจินตทัศน์ข้อมูลมีหลายวิธี มีเครื่องมือและแผนภูมิใหม่ ๆ ที่พัฒนาขึ้นมาอย่างต่อเนื่อง และแต่ละประเภทพยายามสร้างแผนภูมิที่น่าสนใจมากกว่าเดิม สิ่งที่สำคัญการจินตทัศน์ข้อมูลไม่ควรก่อให้เกิดความสับสนและไม่นำเสนอข้อมูลที่ไม่จำเป็น

เทคโนโลยีที่นำมาประยุกต์ใช้ในการจินตทัศน์ข้อมูล

เทคโนโลยีที่นำมาประยุกต์ใช้ในการจินตทัศน์ข้อมูล ปัจจุบันมีเทคโนโลยีหรือซอฟต์แวร์ที่ค่อนข้างหลากหลายในที่นี้ขอยกตัวอย่างเทคโนโลยีบางส่วนที่เป็นที่นิยมใช้ ดังต่อไปนี้

1. Tableau เป็นซอฟต์แวร์ที่สามารถนำข้อมูลจำนวนมากที่มีหลากหลายในองค์กรมาทำการวิเคราะห์เพื่อเป็นข้อมูลเชิงธุรกิจช่วยให้ผู้บริหารมีข้อมูลเชิงลึกเพิ่มความสามารถในการตัดสินใจอย่างรวดเร็วและชาญฉลาดในรูปแบบของจินตทัศน์ข้อมูล (Data Visualization) ซึ่งเป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ไม่ว่าจะเป็นตัวเลข แผนภูมิ กราฟ และอื่น ๆ (“คู่มือการฝึกอบรม,” 2562, หน้า 3)

2. Power BI Desktop คือโปรแกรมที่ใช้ช่วยวิเคราะห์ สรุปผลข้อมูลจำนวนไม่จำกัด จากหลาย ๆ แหล่งข้อมูล ไม่ว่าจะเป็น Excel File, Microsoft Access Database, SQL Server, Oracle, Power Platform เป็นต้น ได้อย่างรวดเร็ว โดยผู้ใช้ไม่ต้องมีความรู้เชิงเทคนิคมากนัก พร้อมสามารถแสดงผลได้ทั้งรูปแบบ ตาราง สรุปผลด้วย Visualization กราฟในรูปแบบต่าง ๆ และยังติดตั้งเพิ่มเติมได้จาก Marketplace และไฮไลท์ก็คือ แสดงผลแบบเป็นแผนที่ได้อีกด้วย โดยสามารถแสดงผลผ่านเว็บไซต์ และอุปกรณ์ Mobile และ Tablet ได้อีกด้วย ทำให้เราได้ข้อมูลที่ถูกต้องและรวดเร็ว พร้อมในการกำหนดกลยุทธ์ และตัดสินใจได้อย่างถูกต้องและแม่นยำในทางธุรกิจ

3. Plotly เป็นเครื่องมือในการสร้างแผนภูมิที่มีความสามารถในการแสดงผลที่โต้ตอบกับผู้ใช้งานได้หลากหลาย มีแผนภูมิมากกว่า 40 ประเภท แสดงผลในรูปแบบของภาพนิทัศน์ (Visualization) เช่น แผนภูมิแท่ง (Bar Chart), แผนภูมิเส้น (Line Chart), แผนภูมิจุด (Scatter Plot) เป็นต้น สามารถนำแผนภูมิที่ทำเสร็จสิ้นแล้วไปเพิ่มลงในเว็บไซต์ที่ต้องการได้ง่ายด้วย Python Framework ที่ชื่อว่า “Plotly Dash” และที่สำคัญสามารถนำไปใช้งาน แก้ไข และเผยแพร่ได้อย่างเสรี (Open Source Library) (ธนกร ทำอิ่นแก้ว, 2564)

4. Gephi เป็นซอฟต์แวร์โอเพ่นซอร์สสำหรับการวิเคราะห์กราฟและเครือข่าย ใช้การเรนเดอร์ 3D เพื่อแสดงเครือข่ายที่มีขนาดใหญ่แบบเรียลไทม์และทำให้การสำรวจมีความรวดเร็ว สถาปัตยกรรมที่ยืดหยุ่นและทำงานได้หลากหลาย ช่วยในการจัดการข้อมูลที่ซับซ้อนและสร้างผลลัพธ์ที่เป็นภาพที่มีคุณค่า Gephi มีคุณสมบัติในการนำเสนอที่หลากหลายในการสำรวจเชิงโต้ตอบและการแปลความหมายของเครือข่าย ให้สามารถเข้าถึง

ข้อมูลเครือข่ายได้ง่าย และสามารถกำหนดพื้นที่ได้ การกรอง การนำทาง การจัดการ และการรวมกลุ่ม Gephi เน้นการแสดงภาพเครือข่ายแบบไดนามิก (Bastian, Heymann, & Jacomy, 2009)

เทคโนโลยีที่นำมาประยุกต์ใช้ในการจินตทัศน์ข้อมูลดังกล่าวเป็นเครื่องมือที่ช่วยให้ผู้ปฏิบัติงานสามารถทำงานได้สะดวกรวดเร็วและง่ายขึ้น ปัจจุบันมีเทคโนโลยีในการจินตทัศน์ข้อมูลที่หลากหลายและมีการพัฒนาซอฟต์แวร์ใหม่ ๆ ออกมาอย่างต่อเนื่อง ซึ่งในบทความนี้นำเสนอเพียงบางส่วนเท่านั้น สำหรับผู้ที่สนใจสามารถศึกษาข้อมูลเพิ่มเติมเพื่อเลือกซอฟต์แวร์ที่เหมาะสมกับข้อมูลที่ต้องการพัฒนาและตรงตามความต้องการกับผู้ใช้งานได้อย่างหลากหลาย

สรุป

บทความนี้มีวัตถุประสงค์เพื่อถ่ายทอดองค์ความรู้เกี่ยวกับจินตทัศน์ข้อมูล (Data Visualization) และเพื่อสร้างความเข้าใจเกี่ยวกับการจินตทัศน์ข้อมูลได้ดียิ่งขึ้น เนื้อหาในบทความนำเสนอความรู้เกี่ยวกับความหมาย การนำเสนอข้อมูลหรือสารสนเทศด้วยภาพหรือภาพกราฟิก โดยใช้สี รูปร่างหรือรูปทรงสัญลักษณ์ และใช้กระบวนการทางเทคนิคในการแปลงข้อมูลหรือสารสนเทศให้กลายเป็นภาพหรือภาพกราฟิก เพื่อช่วยในการมองเห็น ทำให้เข้าใจข้อมูลที่มีความซับซ้อนได้ง่ายขึ้น การนำเสนอส่วนใหญ่จะนำเสนอเป็นแผนภาพ แผนภูมิ แผนผัง ภาพกราฟิก แผนที่ หรืออยู่ในรูปแบบอื่น ๆ ซึ่งนำมาช่วยในการตัดสินใจ เล่าเรื่อง และวิเคราะห์แนวโน้มต่าง ๆ เป็นต้น รูปแบบความสัมพันธ์ของข้อมูลมีหลากหลาย เช่น การจัดอันดับ การเปรียบเทียบ การหักเหความสัมพันธ์ จัดลำดับตามเวลา และการกระจาย เป็นต้น ประเภทของแผนภูมิที่นำใช้ในการจินตทัศน์ข้อมูล เช่น แผนภูมิแท่ง แผนภูมิวงกลม แผนภูมิโดนัท แผนภูมิเส้น และแผนภูมิพื้นที่ เป็นต้น ซึ่งเป็นรูปแบบแผนภูมิที่นิยมนำมาใช้โดยทั่วไป และในปัจจุบันรูปแบบแผนภูมิก็มีหลากหลายมากขึ้น สำหรับกระบวนการจินตทัศน์ข้อมูลมีเป้าหมายเพื่อลดความซับซ้อนและแสดงความสัมพันธ์ของข้อมูลซึ่งนำเสนอไว้สองกระบวนการพื้นฐานให้สามารถนำไปประยุกต์ใช้ และเทคโนโลยีที่นำมาประยุกต์ใช้ในการจินตทัศน์ข้อมูลเพื่อให้การทำงานสะดวกรวดเร็วและง่ายขึ้น ผู้ที่สนใจสามารถนำความรู้ที่ได้ไปต่อยอดเพื่อประโยชน์ทางด้านธุรกิจ ด้านการศึกษา ด้านการวิจัย และอื่น ๆ ต่อไป และควรศึกษากระบวนการและวิธีการเชิงลึกเพื่อเป็นแนวทางในการปฏิบัติได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

เอกสารอ้างอิง

การทำความสะอาดข้อมูล Data cleansing หรือ data cleaning คืออะไร. (2563). สืบค้นจาก https://riccosmartdata.com/data-cleansing-or-data-cleaning/

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล. (2564). สืบค้นจาก https://www.9experttraining.com/articles/การเลือกรูปแบบ-visualization-ให้เหมาะสมกับข้อมูล

การใช้งานแผนภูมิเส้น (Line chart) ให้เหมาะสมกับข้อมูล ใน Google data studio. (2563). สืบค้นจาก https://www.mindphp.com/บทความ/google-for-work/274-google-data-studio/7792-line-chart.html

คู่มือการฝึกอบรม Tableau. (2562). สืบค้นจาก https://www.diw.go.th/km/tableau-manual.pdf

จรัล ทรัพย์เสรี. (2553). รู้จัก Box plot. **For Quality, 16**(148), 35-37.

ณัฐดนัย หวังพระธรรม. (2562). **7 การแสดงการแจกแจง: ฮิสโตแกรมและ Density plots.** สืบค้นจาก https://nutdnuy.medium.com/7-การแสดงการแจกแจง-ฮิสโตแกรมและ-density-plots-e642989c123b

ณัฐวิภา. (2559, 29 กรกฎาคม). เครื่องมือคุณภาพ 7 ชนิด (7 QC Tools) [เว็บบอร์ด]. สืบค้นจาก http://econs.co.th/index.php/2016/07/29/7-qc-tools/

ธนกร ทำอิ่นแก้ว. (2564). **เรียนรู้วิธีการสร้าง Interactive visualization ด้วย plotly.** สืบค้นจาก https://bigdata.go.th/big-data-101/interactive-visualization-with-plotly/

ยุรนันท์ จามจุรี. (2563). **เรียนรู้การสร้างกราฟ Time series และ charts แบบต่าง ๆ ด้วย R studio.** สืบค้นจาก https://bigdata.go.th/big-data-101/covid-19-plot-r-studio/

วรวิกา วัฒนสุนทร และชุติมณฑน์ รักนะ. (2560). การออกแบบจินตทัศน์และการพัฒนาระบบจำลองสำหรับควบคุมสัญญาณไฟจราจร. **วารสารวิชาการสมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย ฉบับวิทยาศาสตร์และเทคโนโลยี, 6**(2), 85-97.

2018 Holiday survey of consumers. (2020). Retrieved from https://vizzlo.com/create/pictograph

Add your data easily. (2021). Retrieved from https://vizzlo.com/create/pie-chart

Ali, S. M., Gupta, N., Nayak, G. K., & Lenka, R. K. (2016). Big data visualization: Tools and challenges. **Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), India,** 656-660. doi:10.1109/IC3I.2016.7918044

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. **Proceedings of the Third International AAAI Conference on Weblogs and Social Media, USA, 3**, 361-362. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/13937

Californian sunlight and visitors to SeaWorld. (2013). Retrieved from https://blog.plotly.com/post/146891210217/spurious-correlations

Chertchom, P. (2017). Output based upon input, the result from data analytics and visualization of TNI registration system's data. **TNI Journal of Business Administration and Languages, 5**(1), 49-54.

Data visualization best-practices for comparison trends. (2020). Retrieved from https://www.lintao-dashboards.com/data-visualizations/best-practices-for-comparison-trends/

Data visualization: What it is and why it matters. (2021). Retrieved from https://www.sas.com/en_us/insights/big-data/data-visualization.html

Forecast of the share of smart connected major home appliances shipments. (2020). Retrieved from https://vizzlo.com/create/donut-chart

Friendly, M. (2008). A brief history of data visualization. In C. H. Chen, W. K. Härdle, & A. Unwin (Eds.), **Handbook of data visualization** (pp. 15-56). doi:10.1007/978-3-540-33037-0_2

Germans are world travel champs. (2020). Retrieved from https://vizzlo.com/create/pictograph

Gordon.silvermanaz. (2014). **What did Singapore export in 2012?** Retrieved from https://en.wikipedia.org/wiki/Treemapping#/media/File:2012_Singapore_Products_Export_Treemap.png

Hot summer days in Leipzig, Germany. (2020). Retrieved from https://vizzlo.com/create/area-chart

Hubspot & Visage. (n.d.). **Data visualization 101: How to design charts and graphs.** Retrieved from https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf

Joshi, Y. K., Chawla, U., & Shukla, S. (2020). Rainfall prediction using data visualisation techniques. **Proceedings of the Confluence 2020 10th International Conference on Cloud Computing, Data Science & Engineering, India,** 327-331. doi:10.1109/Confluence47617.2020.9057928

Matias, M. (n.d.). **Visualize it!: A comprehensive guide to data visualization.** Retrieved from https://www.melissamatias.com/s/ebook-data-visualization-EN.pdf

Nediger, M. (2020). **What is data visualization? (definition, examples, best practices).** Retrieved from https://venngage.com/blog/data-visualization/#1

Netflix: Share of global audience. (2021). Retrieved from https://vizzlo.com/create/stacked-bar-chart

Nowicki, H., & Merenstein, C. (2016). **Radar chart.** Retrieved from https://www.cs.middlebury.edu/~candrews/showcase/infovis_techniques_s16/radar_chart/

Philp, R. (2020). **My favorite tools: Alberto Cairo on data visualization.** Retrieved from https://gijn.org/2020/11/24/my-favorite-tools-alberto-cairo-on-data-visualization/

Rana, A. (2020, January 3). Most populated countries in the future 2020-2100 [Video file]. Retrieved from https://www.youtube.com/watch?v=VPTkNGeMeis

Roshdy, A., Sharaf, N., Saad, M., & Abdennadher, S. (2018). Generic data visualization platform. **Proceedings of the 2018 22nd International Conference Information Visualisation (IV), Italy,** 56-57. doi:10.1109/iV.2018.00020

Scott, D. W. (2008). Histograms: Theory and practice. In **Multivariate density estimation: Theory, practice, and visualization** (pp. 47-94). doi:10.1002/9780470316849.ch3

Strecker, J. (2012). **Data visualization in review: Part of the strategic evaluation on communicating research for influence.** Retrieved from http://hdl.handle.net/10625/49286

Susie. (2020, April 6). K-means centroid deviation [Web log post]. Retrieved from https://bl.ocks.org/susielu/c9170f67e06af8adfbe29ba7617c2cfa

Temperature over seasons in Germany. (2021). Retrieved from https://vizzlo.com/create/nightingales-rose-chart

Where our energy comes from. (2021). Retrieved from https://vizzlo.com/create/stacked-bar-chart

Wilke, C. O. (2019). **Fundamentals of data visualization: A primer on making informative and compelling figures.** Retrieved from https://clauswilke.com/dataviz/aesthetic-mapping.html

Zifan, A. (2017). **Florida counties: United States presidential election, 2016.** Retrieved from https://en.wikipedia.org/wiki/Treemapping#/media/File:United_States_presidential_election_in_Florida,_2016.svg