

รายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าอบรม สัมมนา หรือประชุมวิชาการ

ข้าพเจ้า นายทวีศักดิ์ จันทร์งาม ตำแหน่ง อาจารย์ สังกัด สาขาวิชาสถิติ คณะวิทยาศาสตร์ ขอ
นำเสนอรายงานสรุปเนื้อหาและการนำไปใช้ประโยชน์จากการเข้าร่วมอบรมเชิงปฏิบัติการภายใต้หัวข้อ
Workshop on Using R for Analysis of Big Data เมื่อวันที่ 8 - 11 สิงหาคม พ.ศ. 2559 จัดโดยภาควิชา
คณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ตามหนังสือขออนุญาต เลขที่
ศธ.0523.4.7/283 ลงวันที่ 14 กรกฎาคม พ.ศ. 2559 ซึ่งการเข้าร่วมอบรมเชิงปฏิบัติการดังกล่าวข้าพเจ้าได้
เลือกใช้งบประมาณการพัฒนาบุคลากรตามกรณีที่ 4 เข้าอบรมเชิงปฏิบัติการ ดังนั้นจึงขอนำเสนอสรุปเนื้อหา
และการนำไปใช้ประโยชน์ของการอบรมเชิงปฏิบัติการ ดังต่อไปนี้

จากการที่เข้าร่วมการอบรมเชิงปฏิบัติการภายใต้หัวข้อ Workshop on Using R for Analysis of
Big Data เมื่อวันที่ 8 - 11 สิงหาคม พ.ศ. 2559 ณ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัย
เทคโนโลยีพระจอมเกล้าธนบุรี สามารถสรุปเป็นด้าน ๆ ดังนี้

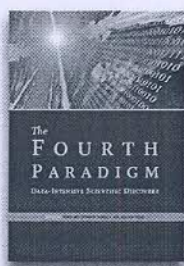
ด้านเนื้อหา

เนื้อหาที่เข้าร่วมอบรมเชิงปฏิบัติการโดย Professor Nagaray, Dr. George Ostrouchov และ
Dr. Andrew M. Raim แยกเนื้อหาการอบรมเชิงปฏิบัติการได้ดังรายละเอียดต่อไปนี้

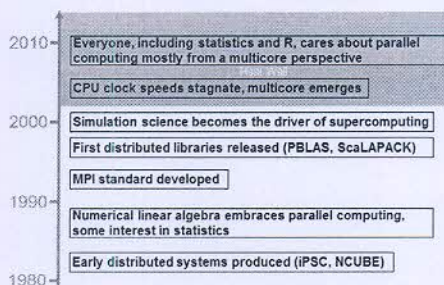
Introduction to Big Data and Why R?

วิทยากรได้อธิบายถึงความรู้เบื้องต้นเกี่ยวกับข้อมูลขนาดใหญ่ (Big Data) โดยกล่าวถึงวิวัฒนาการของ
การเกิดขึ้นของข้อมูลขนาดใหญ่ อาทิ ข้อมูลที่ได้จากเซนเซอร์ (sensors) ของดาวเทียม ข้อมูลเกี่ยวกับชั้น
บรรยากาศ ข้อมูลที่ได้จากกล้องจุลทรรศน์ (microscopy) ข้อมูลนิวตรอน (neutron science) ข้อมูลทางด้าน
ธุรกิจ ข้อมูลทะเบียนราษฎร์ของรัฐบาล ข้อมูลภาษี ข้อมูลจากอินเทอร์เน็ต เป็นต้น จากนั้นวิทยากรได้ชี้ถึง
ความสำคัญของข้อมูลว่าเป็นรากฐานที่สำคัญวิทยาศาสตร์และวิวัฒนาการของนักสถิติในการจัดการและ
วิเคราะห์ข้อมูล เป็นต้น

Data may be the "fourth pillar" of science



Parallel Computing from a Statistician's Perspective



Introduction to R

วิทยากรได้อธิบายถึงความรู้เบื้องต้นเกี่ยวกับโปรแกรมอาร์ (R-programming) ซึ่งเป็นซอฟต์แวร์ที่
อนุญาตให้ใช้ได้โดยไม่ต้องเสียค่าใช้จ่ายใด ๆ ภายใต้ลิขสิทธิ์แบบ GNU General Public License ในรูปรหัส
source code โดยโปรแกรมอาร์ ยังเป็นซอฟต์แวร์ที่รวมเอาคุณสมบัติด้านการจัดการข้อมูล การคำนวณ และ
การแสดงผลทางกราฟิกไว้ด้วยกันอย่างดี โดยวิทยากรได้อธิบายถึงสภาพแวดล้อม (Environment) ของโปรแกรม
อาร์ว่ามีอะไรบ้าง อาทิ มุมมองของโปรแกรมที่เรียกว่า R-GUI โปรแกรมมีสภาพแวดล้อมแบบ object

oriented ซึ่งหมายถึงสิ่งต่าง ๆ ที่อยู่หน่วยความจำของโปรแกรมอาร์ จะเป็นวัตถุที่จะต้องมีการมีชื่อกำกับเสมอ อีกทั้งโปรแกรมอาร์ยังมีข้อดีที่เป็นจุดเด่น คือ มีโปรแกรมประยุกต์ (package) ที่ผู้ใช้โปรแกรมอาร์ทั่วโลกเขียนขึ้นมาเพื่อใช้เฉพาะเรื่องเป็นจำนวนมาก ผู้ใช้อื่น ๆ สามารถนำโปรแกรมประยุกต์ดังกล่าวมาใช้ได้เลยโดยไม่ต้องกังวลเรื่องลิขสิทธิ์ จากนั้นวิทยากรได้อธิบายถึงคำสั่งเบื้องต้นต่าง ๆ ที่ใช้ในโปรแกรมอาร์ อาทิ คำสั่ง help รูปแบบคำสั่งการกำหนดวัตถุ คำสั่งการคำนวณต่าง ๆ คำสั่งเมทริกซ์ (Matrices) คำสั่งเวกเตอร์ (vectors) คำสั่งอาร์เรย์ (array) คำสั่งเงื่อนไขการเปรียบเทียบ เช่น if then else เป็นต้น นอกจากนั้นวิทยากรได้แนะนำโปรแกรมเพิ่มเติมอื่น ๆ ที่ใช้ร่วมกับโปรแกรมอาร์ โดยจะทำให้การทำงานของโปรแกรมอาร์มีประสิทธิภาพมากยิ่งขึ้น เช่น R Studio ซึ่งเป็นโปรแกรมในกลุ่ม Integrated Development Environment (IDE) ที่นำมาจัดรูปแบบมุมมองของโปรแกรมอาร์ให้สามารถใช้งานได้ง่ายขึ้น รวมถึงยังโปรแกรมประยุกต์ที่สำคัญที่จะอำนวยความสะดวกในการจัดการและวิเคราะห์ข้อมูล อาทิ R Commander ที่อำนวยความสะดวกโดยการสร้างสภาพแวดล้อมแบบเมนูและเลือกใช้, Rattle ที่อำนวยความสะดวกในการจัดการเหมืองข้อมูล (data mining) จากนั้นวิทยากรได้ให้ผู้เข้าร่วมการอบรมได้ใช้คำสั่งต่าง ๆ ในการจัดการและวิเคราะห์ข้อมูล เช่น การแจกแจงข้อมูล (distributions) การพรรณนาข้อมูล (descriptive statistics) และการอนุมานเบื้องต้น (basic inference)

Basic Data Analysis

วิทยากรได้อธิบายและนำผู้เข้าร่วมการอบรมใช้โปรแกรมอาร์และโปรแกรมประยุกต์ ในวิเคราะห์ข้อมูลขนาดใหญ่ โดยใช้วิธีการทางสถิติ อาทิ การวิเคราะห์การถดถอย (regression) การวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (general linear model: GLM) การวิเคราะห์การจำแนก (classification) การวิเคราะห์การจำแนกและการถดถอยแบบต้นไม้ (classification and regression trees) การวิเคราะห์ random forest การวิเคราะห์การแบ่งกลุ่ม (clustering) การวิเคราะห์ส่วนประกอบหลัก (principal components analysis) การวิเคราะห์เครือข่าย neural เป็นต้น

Getting your data into R

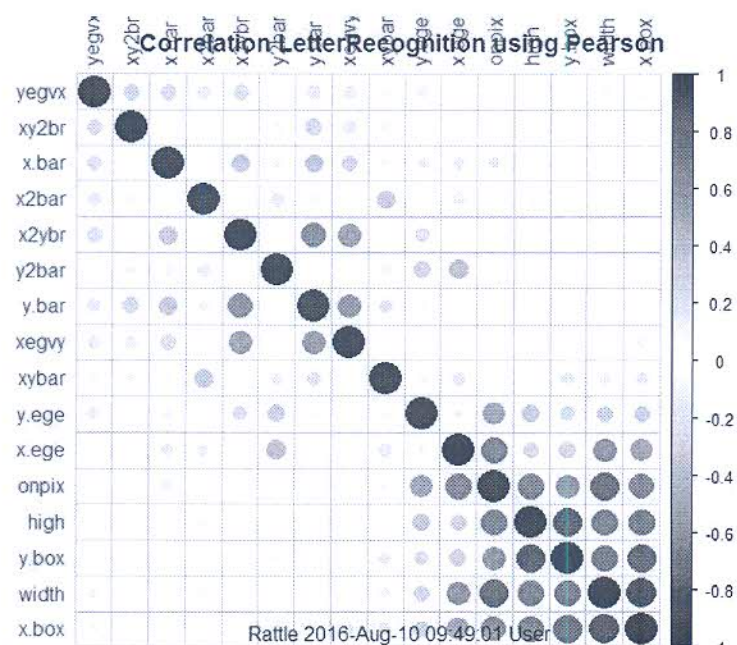
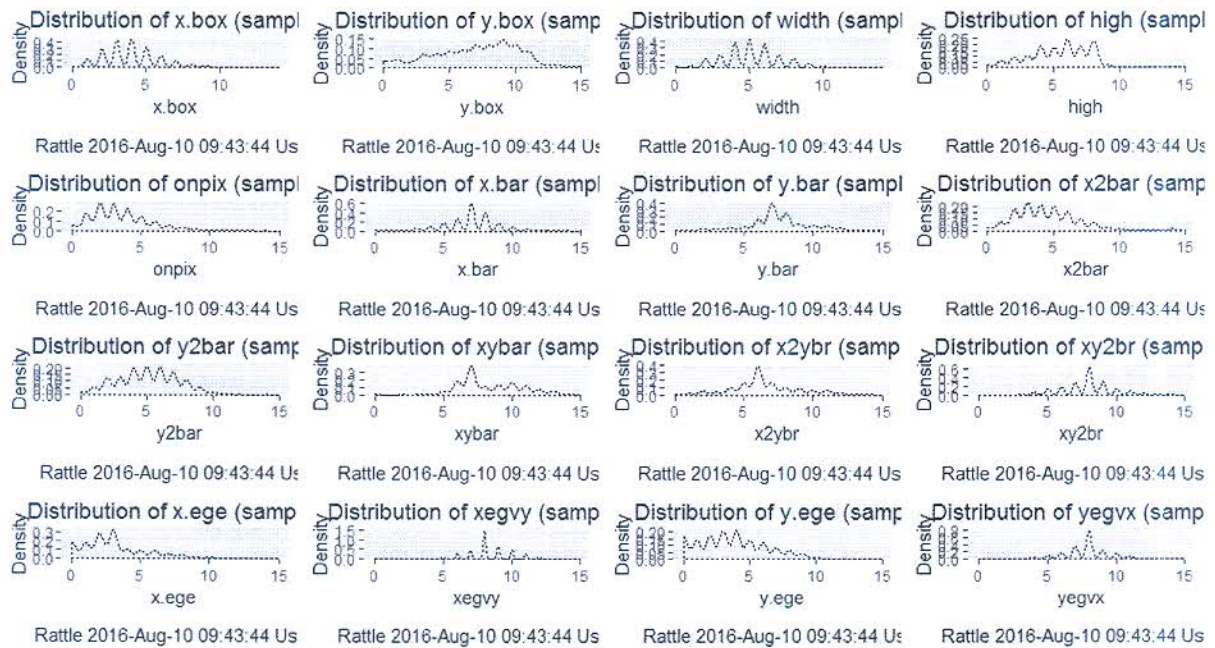
วิทยากรได้อธิบายและนำผู้เข้าร่วมการอบรมใช้โปรแกรมอาร์และโปรแกรมประยุกต์ rattle ในการนำเข้าข้อมูลที่อยู่ในรูปแบบไฟล์ต่าง ๆ อาทิ CSV files, Binary files (rhd5), SQL databases, SAS (sas7bdat) และ Big datasets โดย

Evaluating Variability and Uncertainty

วิทยากรได้อธิบายและนำผู้เข้าร่วมการอบรมใช้โปรแกรมอาร์ในการประเมินความผันแปรและความไม่แน่นอนของข้อมูล ด้วยวิธีการ Bootstrap, Cross-validation โดยการโปรแกรม (writing code)

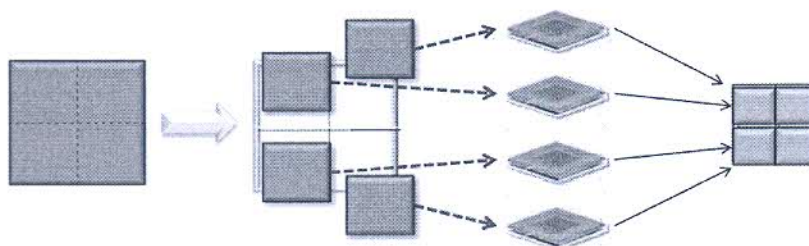
Graphics

วิทยากรได้อธิบายและนำผู้เข้าร่วมการอบรมใช้โปรแกรมอาร์ในการสร้างรูปแบบภาพต่าง ๆ เพื่อใช้ในการวิเคราะห์ข้อมูลโดยใช้ Traditional Graphics, ggplot2 และใช้ knitr ดังรูปภาพตัวอย่าง



Parallels

วิทยาการได้อธิบายถึงหลักการของโปรแกรมแบบขนาน ซึ่งมีจุดประสงค์ทำให้โปรแกรมทำงานได้เร็วขึ้น โดยเฉพาะการวิเคราะห์ข้อมูลขนาดใหญ่ (big data) โดยอยู่บนหลักการที่ว่ามีการแบ่งงานใหญ่เป็นงานย่อย ๆ หลาย ๆ งาน จากนั้นส่งงานย่อยไปทำงานยังเครื่องประมวลผลหลาย ๆ เครื่องพร้อมกัน แล้วจึงทำการรวบรวมผลที่ได้จากเครื่องประมวลผลต่าง ๆ ดังภาพด้านล่าง



โดยวิทยาการได้นำโปรแกรมประยุกต์ที่ใช้ในโปรแกรมอาร์มาใช้ อาทิ pdbR จากนั้นวิทยาการได้สาธิตและให้ผู้เข้าร่วมอบรมวิเคราะห์ข้อมูลแบบขนาน อาทิ Parallel Bootstrap, Parallel Cross-validation และ Parallel randomForest เป็นต้น รวมทั้งการนำเข้าข้อมูลขนาดใหญ่ด้วยการโปรแกรมแบบขนานด้วย อาทิ Parallel CSV file input, Parallel HDF5 file input และ Parallel ADIOs file input รวมถึงการสร้างโปรแกรมคู่ขนานในการจัดการเมทริกซ์ขนาดใหญ่ให้มีความรวดเร็วในการประมวลผลด้วย เป็นต้น

ด้านการนำไปใช้ประโยชน์

จากการเข้าร่วมประชุมอบรมเชิงปฏิบัติการภายใต้หัวข้อ Workshop on Using R for Analysis of Big Data สามารถนำไปใช้ประโยชน์ในด้านต่าง ๆ ดังนี้

ส่วนที่ 1 การเรียนการสอน

สามารถนำความรู้จากการเข้าร่วมการอบรมเชิงปฏิบัติการมาประยุกต์ใช้กับการเรียนการสอนได้มาก ซึ่งสามารถนำมาใช้ในการเรียนการสอนวิชา สถ315 การวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติ ได้เป็นอย่างมากเนื่องจากได้วิธีการในการใช้โปรแกรมทางสถิติเพื่อใช้ในการวิเคราะห์ข้อมูล รวมถึงการแปลผลจากผลลัพธ์ของโปรแกรมทางสถิติดังกล่าว

ส่วนที่ 2 การวิจัย

สามารถนำความรู้ที่ได้จากการเข้าร่วมการอบรมเชิงปฏิบัติการ ไปใช้ประโยชน์ในการทำงานวิจัยด้านการวิเคราะห์ข้อมูลขนาดใหญ่ การนำเสนอรูปภาพที่มีมุมมองหลายด้าน เนื่องจากวิทยาการได้นำเสนอวิธีการใหม่ ๆ และข้อสังเกตต่าง ๆ ในการวิเคราะห์ข้อมูล ซึ่งจะเป็นประโยชน์ในการทำวิจัยเชิงประยุกต์ เป็นต้น

(นายทวีศักดิ์ จันทร์งาม)

26 สิงหาคม 2559

ความคิดเห็นของผู้บังคับบัญชาชั้นต้น (ประธานหลักสูตร/เลขานุการคณะ/หัวหน้างาน)
 ชวิทย์ โปแลนด์ R. ทำตัวได้รับคณบดีมอบหมายในกรณีงานนี้จริง ๆ จังเจ้ตลอดทำทุกอย่างที่สั่งได้
 หน้าที่มอบหมายมาแล้วที่ได้งานมาของม.ให้ใน ม.นี้ที่ทำงานไม่ท้อแท้ไม่ท้อแท้ไม่ท้อแท้
 นื่องหน้าวัน

.....
 (นาย ชวิทย์ โปแลนด์)
 ๒๙ / ๑๑ / ๕๙

ความคิดเห็นของคณบดีคณะวิทยาศาสตร์หรือผู้แทน

.....

.....
 (.....)
/...../.....