**IMDb Movies**



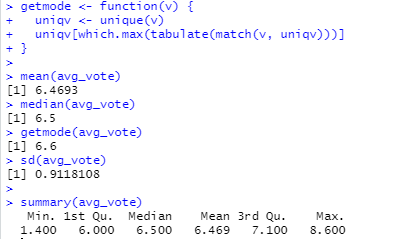
avg\_vote (คะแนนโหวตเฉลี่ย) มีหน่วยเป็น คะแนน

income (รายได้ภาพยนตร์รวมทั่วโลก) มีหน่วยเป็น ล้านดอลลาร์สหรัฐ



**คำนวณหาค่าสถิติพื้นฐาน**

คอลัมน์ avg\_vote (คะแนนโหวตเฉลี่ย)



Mean = 6.4693 คะแนน

Median = 6.5 คะแนน

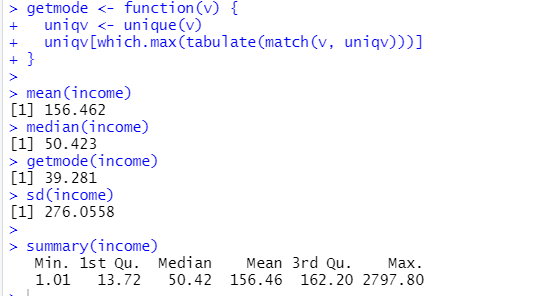
Mode = 6.6 คะแนน

S.D. = 0.9118108

Min = 1.4 คะแนน

Max = 8.6 คะแนน

คอลัมน์ income(รายได้รวมทั่วโลก)



Mean = 156.462 ล้านดอลลาร์สหรัฐ

Median = 50.423 ล้านดอลลาร์สหรัฐ

Mode = 39.281 ล้านดอลลาร์สหรัฐ

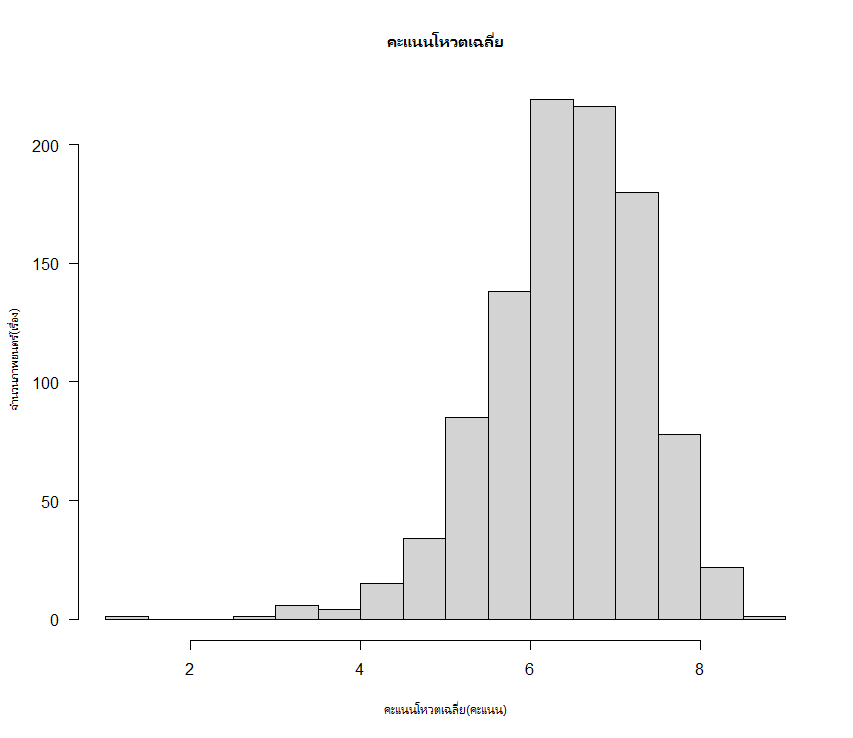
S.D. = 276.0558

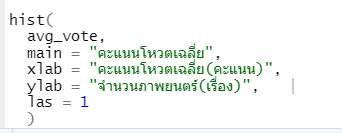
Min = 1.01 ล้านดอลลาร์สหรัฐ

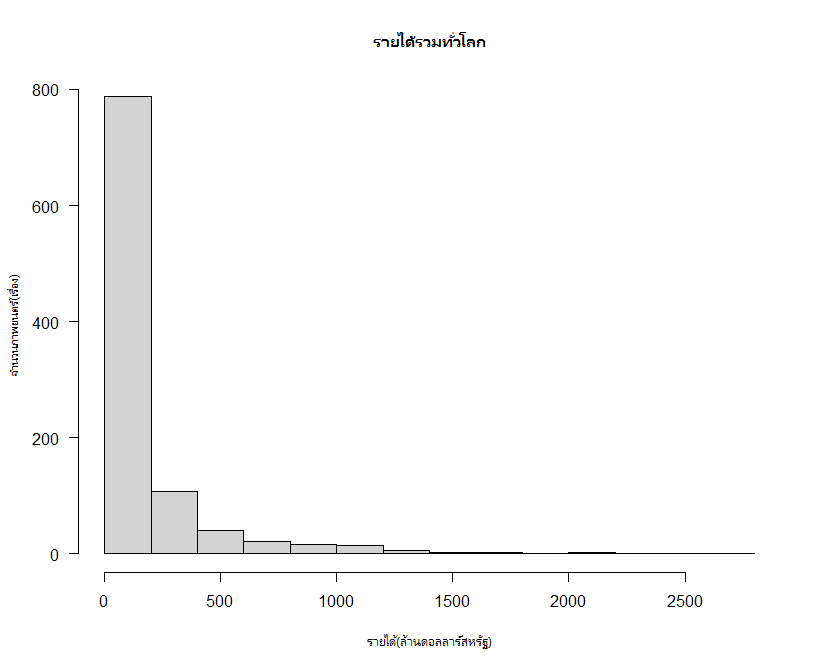
Max = 2797.80 ล้านดอลลาร์สหรัฐ

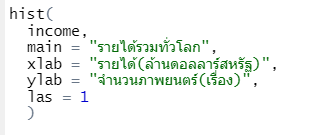
**วาดกราฟ**

**-** Histogram

คอลัมน์ avg\_vote (คะแนนโหวตเฉลี่ย)

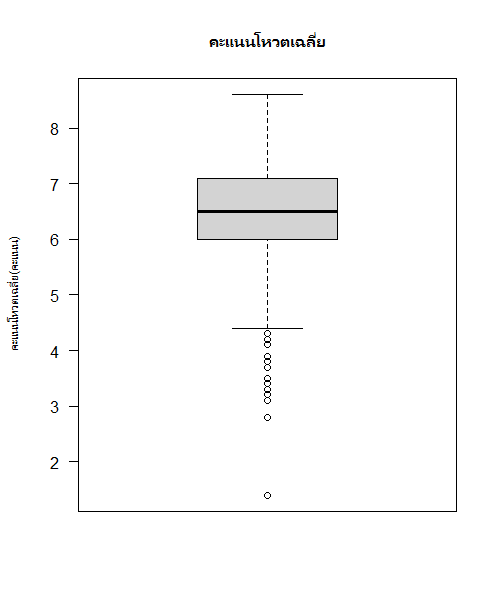


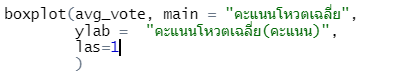
คอลัมน์ income(รายได้รวมทั่วโลก)



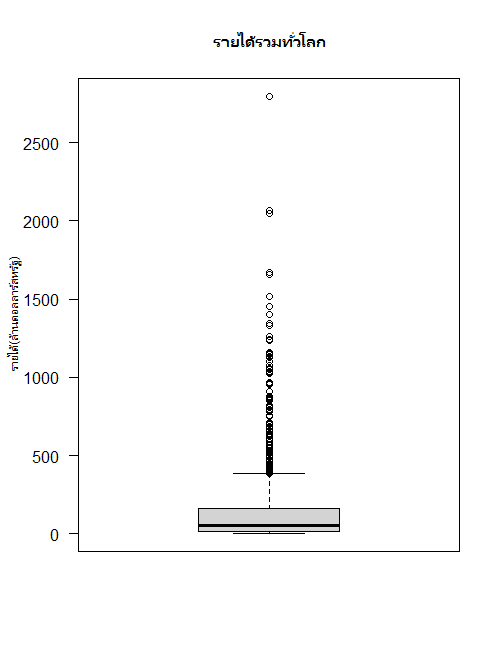
- Box Plot

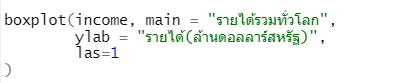
คอลัมน์ avg\_vote (คะแนนโหวตเฉลี่ย)





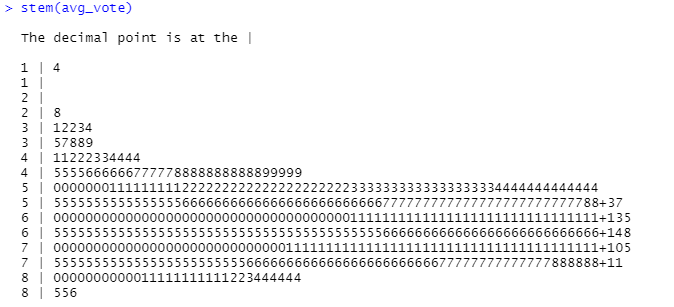
คอลัมน์ income(รายได้รวมทั่วโลก)





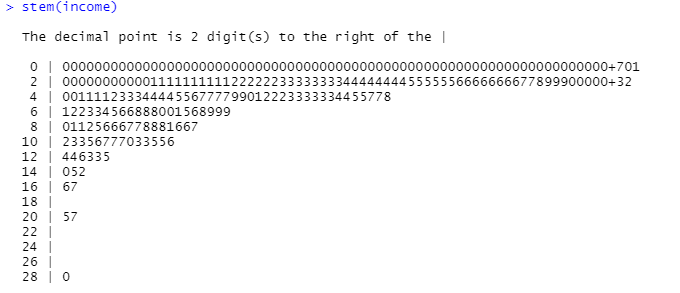
-Stem and Leave

คอลัมน์ avg\_vote (คะแนนโหวตเฉลี่ย)

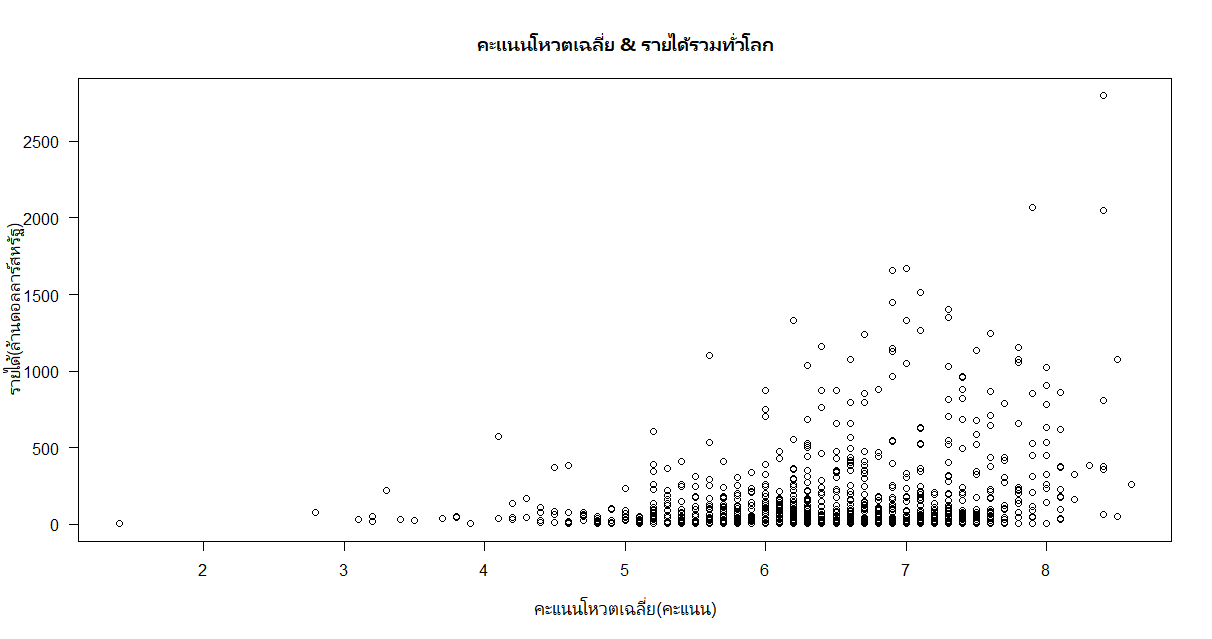


คะแนนโหวตเฉลี่ย(คะแนน)

คอลัมน์ income(รายได้รวมทั่วโลก)



รายได้(ล้านดอลลาร์สหรัฐ)

 -XY (Scatter) Plot

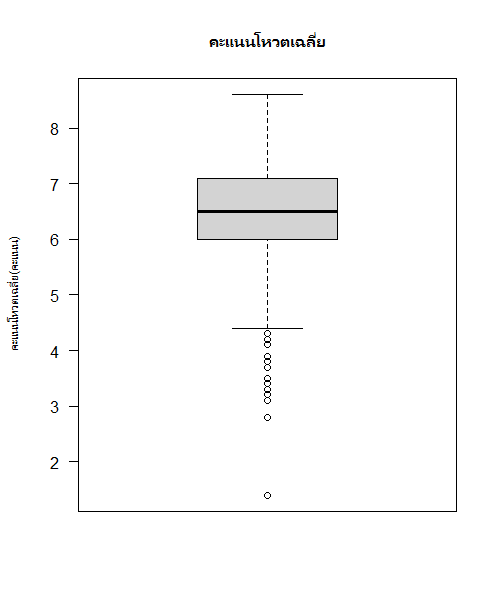
plot(avg\_vote,income)

สาเหตุที่เลือก ตัวแปรต้น (x) เป็น คะแนนโหวตเฉลี่ย และ ตัวแปรตาม (y) เป็น รายได้รวมทั่วโลก เพราะผมอยากรู้ว่ารายได้ภาพยนตร์จะขึ้นอยู่กับคะแนนโหวตหรือไม่

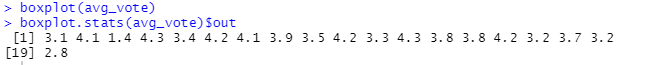
เช่น ถ้าคะแนนเยอะรายได้ก็จะเยอะ คะแนนน้อยรายได้ก็จะน้อย

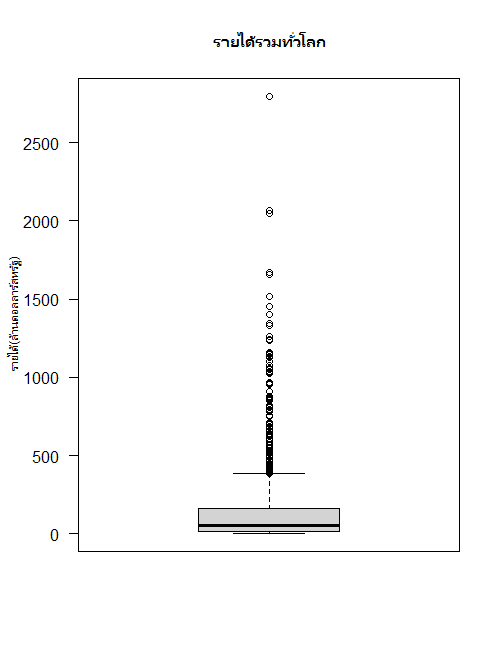
-Outlier

คอลัมน์ avg\_vote (คะแนนโหวตเฉลี่ย)

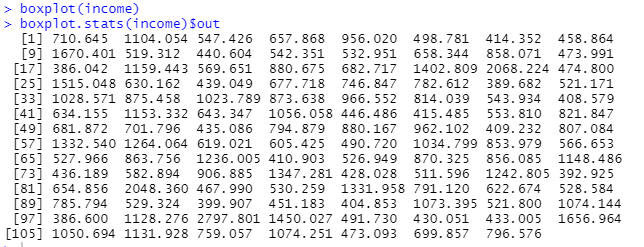


ค่าที่มีแนวโน้มจะเป็น outlier ได้แก่



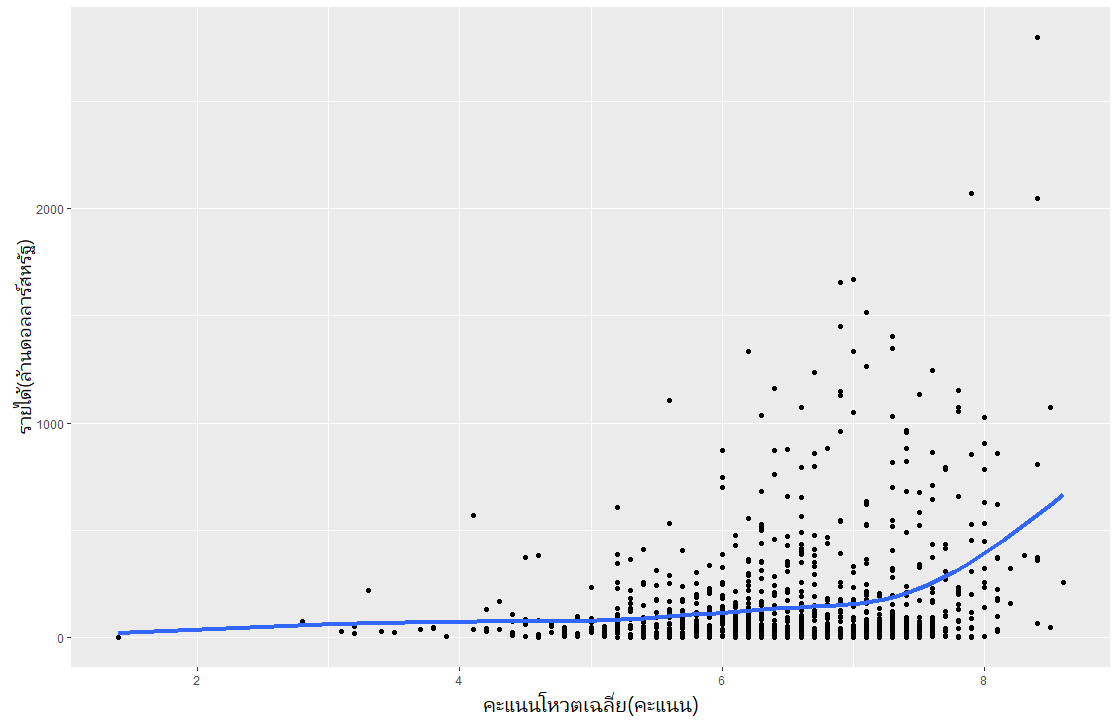
คอลัมน์ income(รายได้รวมทั่วโลก)

ค่าที่มีแนวโน้มจะเป็น outlier ได้แก่



**บทวิเคราะห์ข้อมูลจากกราฟ**

**คะแนนโหวตเฉลี่ยและรายได้รวมทั่วโลกของภาพยนตร์**



จากข้อมูลจากกราฟความสัมพันธ์ระหว่างคะแนนโหวตเฉลี่ยและรายได้ วิเคราะห์ได้ว่า

เมื่อคะแนนโหวตสูง รายได้ของภาพยนตร์ก็จะสูงด้วย ซึ่งผมคิดว่าอาจจะเกิดจากคนดูภาพยนตร์จะดูคะแนนโหวตของภาพยนตร์ก่อนไปดู ถ้าคะแนนเยอะก็จะไปดู ทำให้รายได้ของภาพยนตร์เรื่องนั้นๆสูง ถ้าคะแนนน้อยก็จะไม่ดู ทำให้รายได้ของภาพยนตร์เรื่องนั้นๆต่ำ

**Source Code**



