The Privacy Pickle

Read the paper titled “*The Privacy Pickle”.*Then read the article “Big Data Ethics”. Do you agree with the Predictive Analytics applications mentioned in the paper?  Do they raise ethical concerns?  Why or why not?

My belief is that using data that has not been provided to you is a similar to stealing. Therefore, I do not agree with what HP is doing. It would be okay if HP was transparent about their intentions and had employees sign waivers allowing collection of this kind of information to conduct analysis. However, in the article “Privacy Pickle”, they mention that most of the employees are not aware that this is happening. This sets a dangerous precedent. If you want to use that data, the person must be willing to share that information with you. Facebook currently uses its user’s data, but there is an option if you would like to opt out. Providing people with the choice would lessen the ethical dilemma.

The one point that stuck out to me from the article “Big Data Ethics” is “The law is a powerful element of Big Data Ethics, but it is far from able to handle the many use cases and nuanced scenarios that arise. Organizational principles, institutional statements of ethics, self-policing, and other forms of ethical guidance are also needed.” 1 I agree with this statement, the law is not equipped to handle this kind of nuance. This is very relevant to the discussion that is happening today on whether your Internet Service Provider should be able to collect your internet history and sell it.

Self-policing does not always work, and when it does not, regulations should be in place that protect the privacy of users. Just like the healthcare and food industry is regulated by HIPPA and the FDA, I believe the data collection industry will also give birth to a new regulatory agency in the future.

Those of us enrolled in this program understand the benefits of predictive analytics. I can understand from a business perspective the benefits of being able to collect this information. It would create greater competition, additional business, and additional revenue. However, in this instance I believe the cons outweigh the pros.

1. <http://radar.oreilly.com/2014/03/whats-up-with-big-data-ethics.html>

Week 1 Problem – Analyzing the Databases Number 1

For this week, I decided to try out some functions to better understand how R inputs data from external sources. I went ahead and did problem number 1 from chapter 2, Analyzing the Databases.

From the R code that is listed below, I was able to pull the data from the file and create a histogram outlining the frequency of the number of workers. This distribution shows that for each “SIC Code” the number of workers are fairly low. Almost all SIC codes have 100 people or less working on them. Very few have above that and only a handful have more than 300 people working on them.

I need to learn how to make some aesthetic changes to the histogram, but looking for some guidance on that. See histogram output and R code:

#read file

mfg\_DB <- read.csv(file = "/Users/Zeeshan/Desktop/PREDICT 401/ManufacturingDB.csv", header = TRUE, stringsAsFactors = FALSE)

#check structure of file

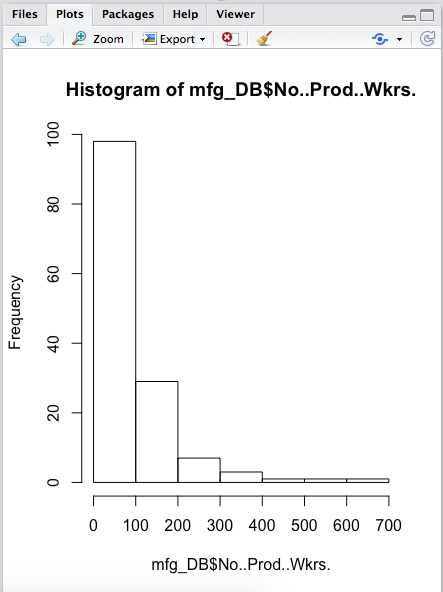
str(mfg\_DB)

#create window

par(mfrow = c(1,1))

#create histogram

hist(mfg\_DB$No..Prod..Wkrs.)



Discussion 1 Follow up – Misuse of Statistics

As we go through the course I’ll be using a lot of NBA examples. I’m a huge fan! The article on the Misuse of Statistics was very insightful. I chose an article about the effects of resting NBA players throughout the season and whether or not it has an effect on the team’s chances of winning.

Gregg Popovich (San Antonio Spurs – Head Coach) is notorious for resting players throughout the season in an attempt to preserve them for the playoffs. His approach, has its critics and defendants, but undoubtedly he’s started a trend. As the article, *NBA Teams Are Resting Players Earlier And Earlier,* states, “Since 2006,1 there have been 609 instances of a player missing a game because he was resting, according to a query of over 50,000 injury reports from Pro Sports Transactions Archive, a database containing NBA transactions dating back to the 1940s.2 Popovich has been responsible for 113 of those 609 games missed by healthy players — more than twice as many as any other coach — and even twice as many as any other franchise.”1 However, this statistic is a little skewed since Gregg Popovich has been one of the longest tenured coaches in the NBA. When comparing to the younger coaches in the league today, they’re just not at the volume that Pop is at.

When looking at the statistics of teams winning and resting players, there is no real correlation between the two, and Whitehead is sure to mention that. “It’s unclear if regular-season rest really can prevent playoff injuries, but that hasn’t stopped Popovich from trying.”1 The last time the Spurs won a championship was in 2014. He used this strategy during that season as well. But, when using the learnings from *Misuse of Statistics* one can argue that the resting of players is not the reason they won. They were simply just the better team. One can argue the reason they did win was due to the emergence of Kawhi Leonard.

The author of this article makes some very good points and is clear when stating that the correlation has not been proven. He simply states the fact that coaches are leaning towards resting players more and earlier in the season. He also provides statistics that show this. Whitehead, does not try to sway the reader in either direction. If I was writing this article, I would use the same approach. However, I did disagree with the statistic about Gregg Popovich players that are healthy missing 113/609 games as mentioned earlier.

1. Whitehead, T. (2017, March 31). NBA Teams Are Resting Players Earlier And Earlier. Retrieved April 1, 2017, from https://fivethirtyeight.com/features/nba-teams-are-resting-players-earlier-and-earlier/

Discussion 2 Comments

Recreation of my histogram

mfg\_DB <- read.csv(file = "/Users/Zeeshan/Desktop/PREDICT 401/Week 1/ManufacturingDB.csv", header = TRUE, stringsAsFactors = FALSE)

#check structure of file

str(mfg\_DB)

#create window

par(mfrow = c(1,1))

#create histogram

hist(mfg\_DB$No..Prod..Wkrs.)

#create histrogram using ggplot2

remove.packages(c("ggplot2", "data.table"))

install.packages('ggplot2', dependencies = TRUE)

library(ggplot2)

# Custom color code

myBlue <- "#377EB8"

# Change the fill color to myBlue

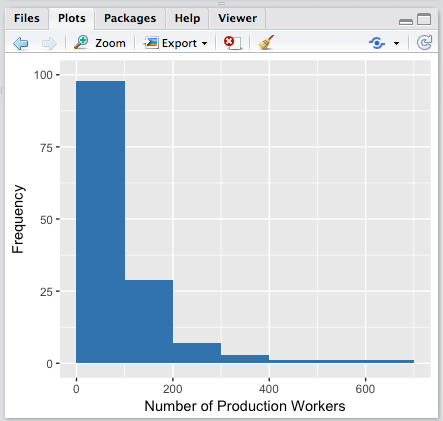
hist <- ggplot(mfg\_DB, aes(x = mfg\_DB$No..Prod..Wkrs.)) +

geom\_histogram(breaks=seq(0, 700, by=100), fill = myBlue) +

xlab("Number of Production Workers") +

ylab("Frequency")

hist + ylim(0,100) + xlim(0,700)



library(ggplot2)

ring\_wt <- c(51,53,41,44,57,47,51,44,57,53,53,44,46,51,52,38,55,49,56,42,52,41,54,53,49,43,50,57,56,52,63,46,50,52,44,46,50,69,42,36,62,43,47,55,57,53,47,58,39,42)

ring\_df <- data.frame(ring\_wt)

myColor <- "#FF6347"

ring\_hist <- ggplot() +

aes(x = ring\_df) +

geom\_histogram(breaks=seq(30, 70, by=10), fill = myColor) +

xlab("Ring Weights") +

ylab("Frequency")

hist + ylim(0,10) + xlim(30,70)

