Based on the results in this chapter, suppose you were asked to summarize what you learned about whether first babies arrive late.

Which summary Statistics would you use if you wanted to get a story on the evening news?

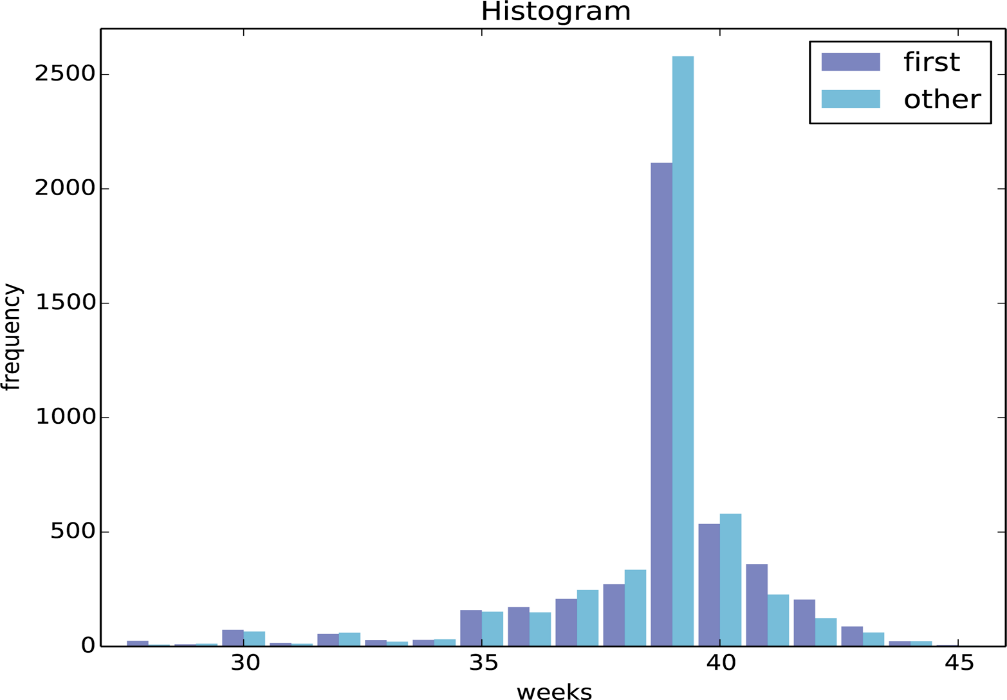
* For this, we would use summary statistics that emphasize differences, using their mean, mode, spread, tails, and outliers as the main descriptive factors.

Which would you use if you wanted to reassure an anxious patient?

* For this, we would choose statistics that put the differences into context, which might consist of mainly the mean, mode, and in some cases, the spread.

“Do First babies arrive late?”

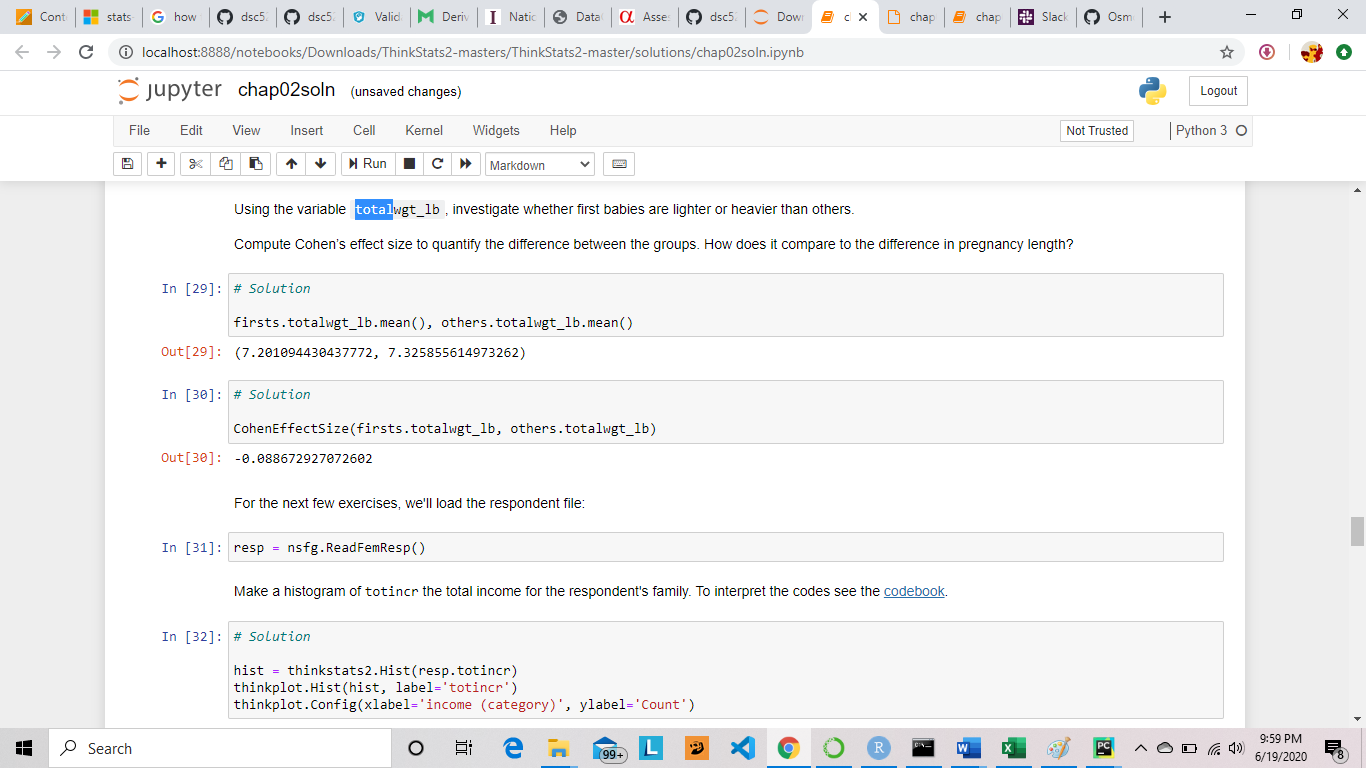
The answer to this question is “it depends”. A lot of questions in statistics can be answered the same way, as the results tend to vary on a case by case basis. Let us observe this histogram for some context:



A full term pregnancy falls in the range of 39-40 weeks, making it “late” after 40 weeks (everything to the right of 40). It can be observed that in pregnancies that lasted 41 weeks, 42 weeks, 43 weeks , and 44 weeks, the number of first babies surpassed the number of other babies, making it a safe conclusion to assume that in this case, first babies do indeed arrive late.

Ex 2-4.

Using the variable totalwgt\_lb, investigate whether first babies are lighter or heavier than others. Compute Cohen’s d to quantify the difference between the groups. How does it compare to the difference in pregnancy length?



By investigating the answer provided from the solutions (because my code would not run for some reason), we can see that the mean weight of the other babies exceeds the mean weight of the first babies, making the first babies lighter than the other babies.

Cohen’s d is an effect size used to indicate the standardized difference between two means.

d =

Cohen’s d is an effective size for the comparison between two means.

d here is approximately = -0.089

A screenshot of a social media post

Description automatically generated

Difference in pregnancy lengths here is approximately = 0.078

Now comparing that to d, it can be observed that both numbers are trivial, as their means do not differ by 0.2 standard deviations.