

Week 3 Quiz

[Help](#)

Warning: You have already made the maximum number of submissions. Additional submissions will not count for credit. You are welcome to try it as a learning exercise.

☐ In accordance with the Coursera Honor Code, I (Sidhartha Sankar Ray) certify that the answers here are my own work.

Question 1

We take a random sample of individuals in a population and identify whether they smoke and if they have cancer. We observe that there is a strong relationship between whether a person in the sample smoked or not and whether they have lung cancer. We claim that the smoking is related to lung cancer in the larger population. We explain we think that the reason for this relationship is because cigarette smoke contains known carcinogens such as arsenic and benzene, which make cells in the lungs become cancerous.

- ☐ This is an example of a predictive data analysis.
- ☐ This is an example of an descriptive data analysis.
- ☐ This is an example of an inferential data analysis.
- ☐ This is an example of a causal data analysis.

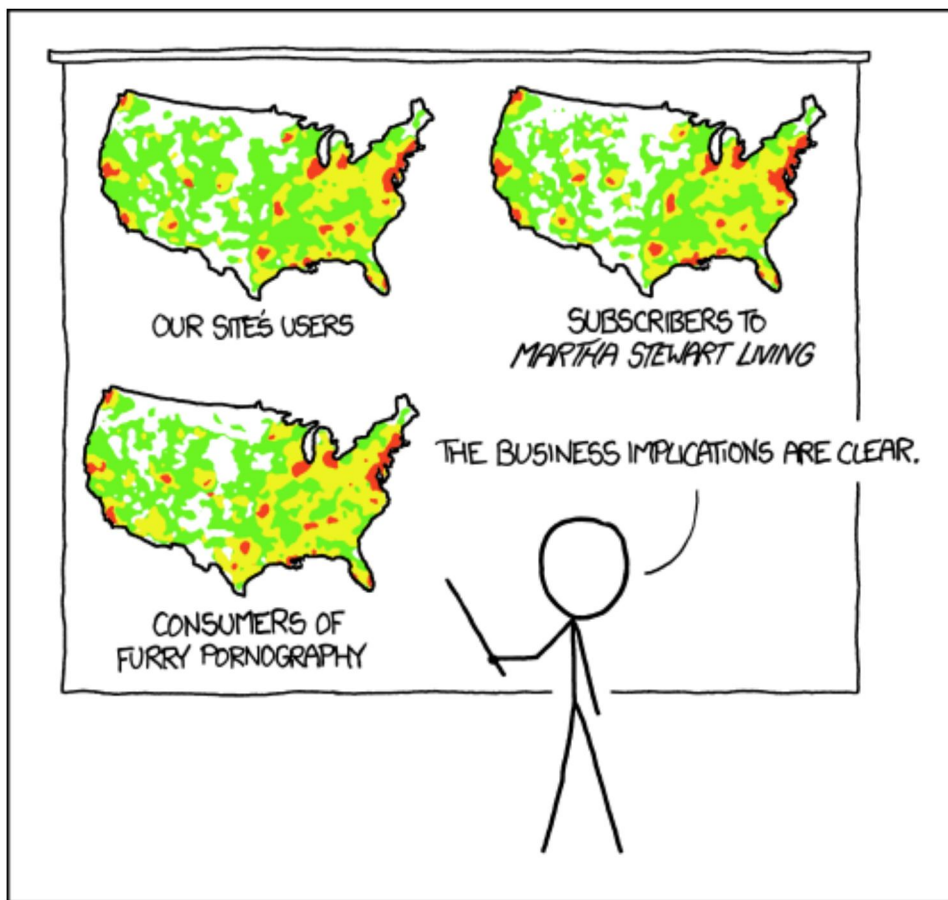
Question 2

What is the most important thing in Data Science?

- ☐ The data.
- ☐ Knowing Hadoop and Pig.
- ☐ Hacking skills.
- ☐ The question you are trying to answer.

Question 3

If the goal of a study was to relate Martha Stewart Living Subscribers to Our Site's Users based on the number of people that lived in each region of the US, what would be the potential problem?



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

[Link to image](http://xkcd.com/1138/)

Source: <http://xkcd.com/1138/>

- ☐ There would be confounding because the number of people that live in an area is related to both Martha Stewart Living Subscribers and Our Site's Users.
- ☐ We would be data dredging because we collected data on thousands of places in the United States.
- ☐ We wouldn't know the sensitivity of our predictions.
- ☐ We couldn't be sure whether subscribing to Martha Stewart Living causes people to be Users of Our Site or the other way around.

Question 4

What is an experimental design tool that can be used to address variables that may be

confounders at the design phase of an experiment?

- ☐ Using regression models.
- ☐ Stratifying variables.
- ☐ Using data from a database.
- ☐ Using all the data you have access too.

Question 5

What is the reason behind the explosion of interest in big data?

- ☐ We recently discovered ways to use data to answer scientific and business questions.
- ☐ Technologies like Hadoop and Map Reduce started the big data era.
- ☐ The price and difficulty of collecting and storing data has dramatically dropped.
- ☐ We have better experimental design now than previously.

☐ In accordance with the Coursera Honor Code, I (Sidhartha Sankar Ray) certify that the answers here are my own work.

Submit Answers

Save Answers

You cannot submit your work until you agree to the Honor Code. Thanks!

