ECE 20875

Python for Data Science

Chris Brinton and David Inouye

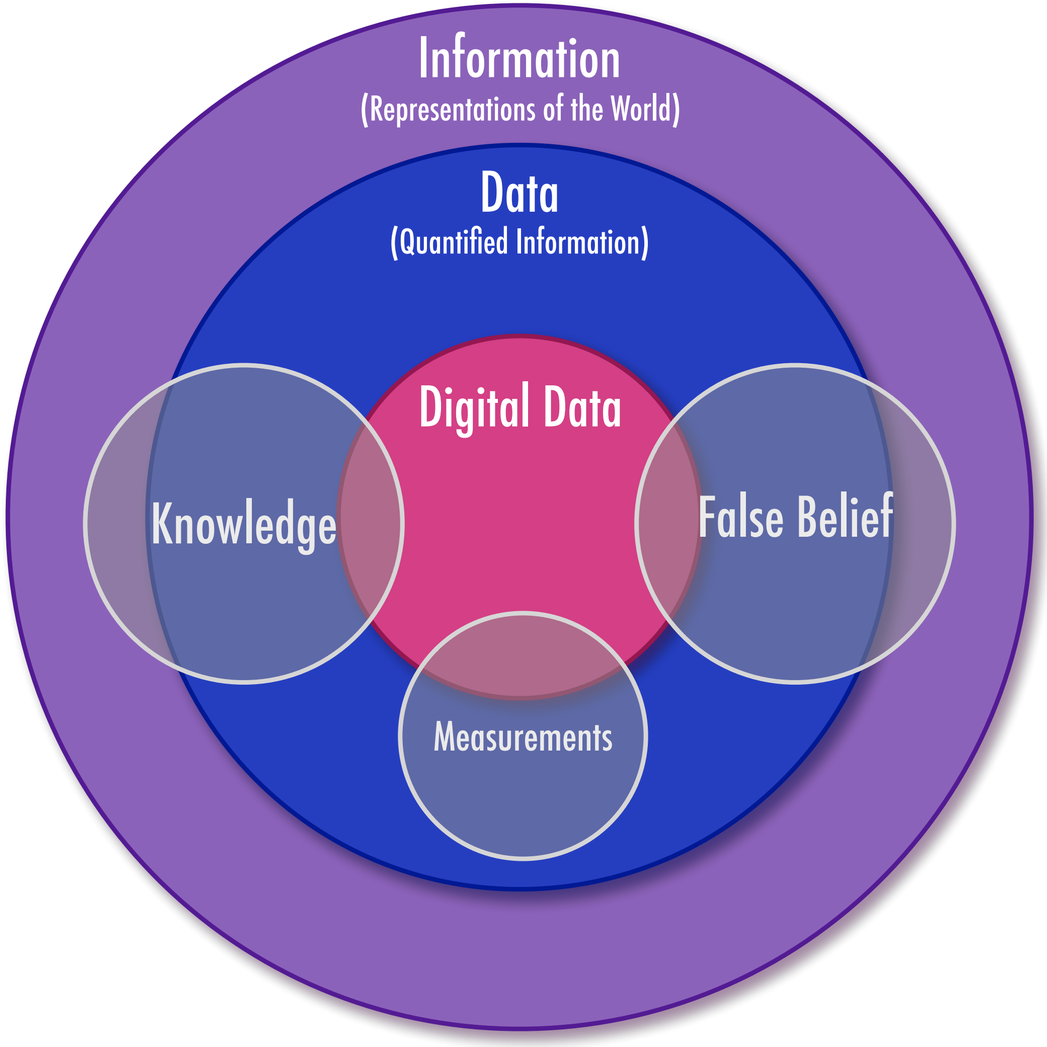
**(Adapted from material developed by
Prof. Milind Kulkarni and Prof. Chris Brinton)**

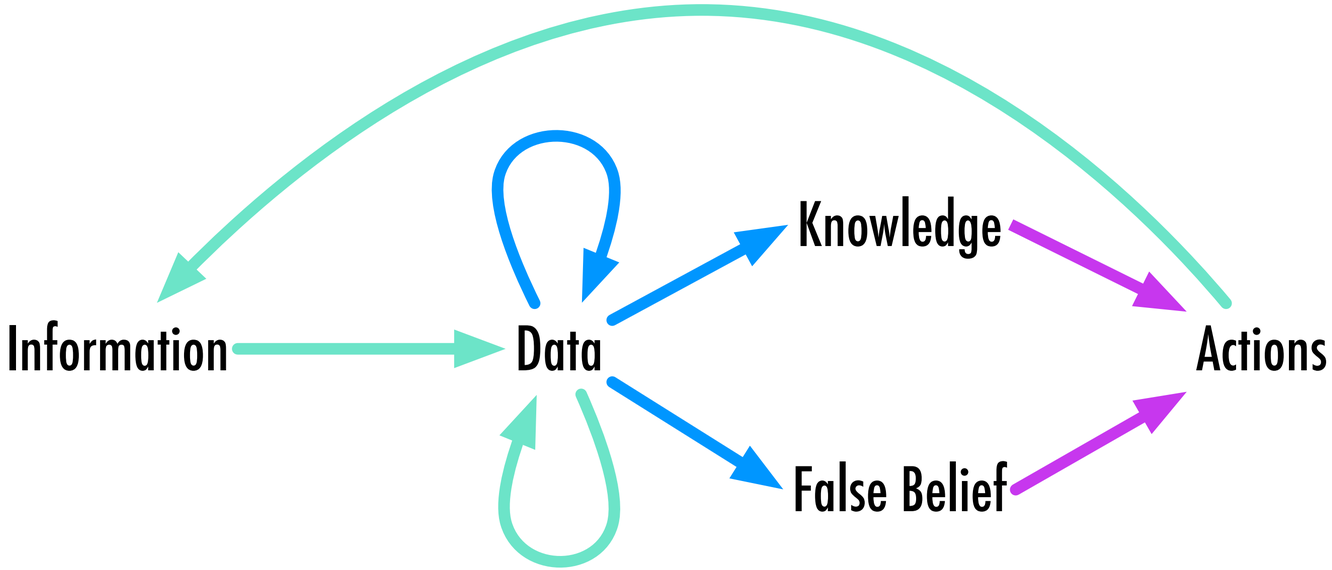
MWF, 12:30pm-1:20pm

Section I: WALC 1055
Section II: FRNY G124

what is data?

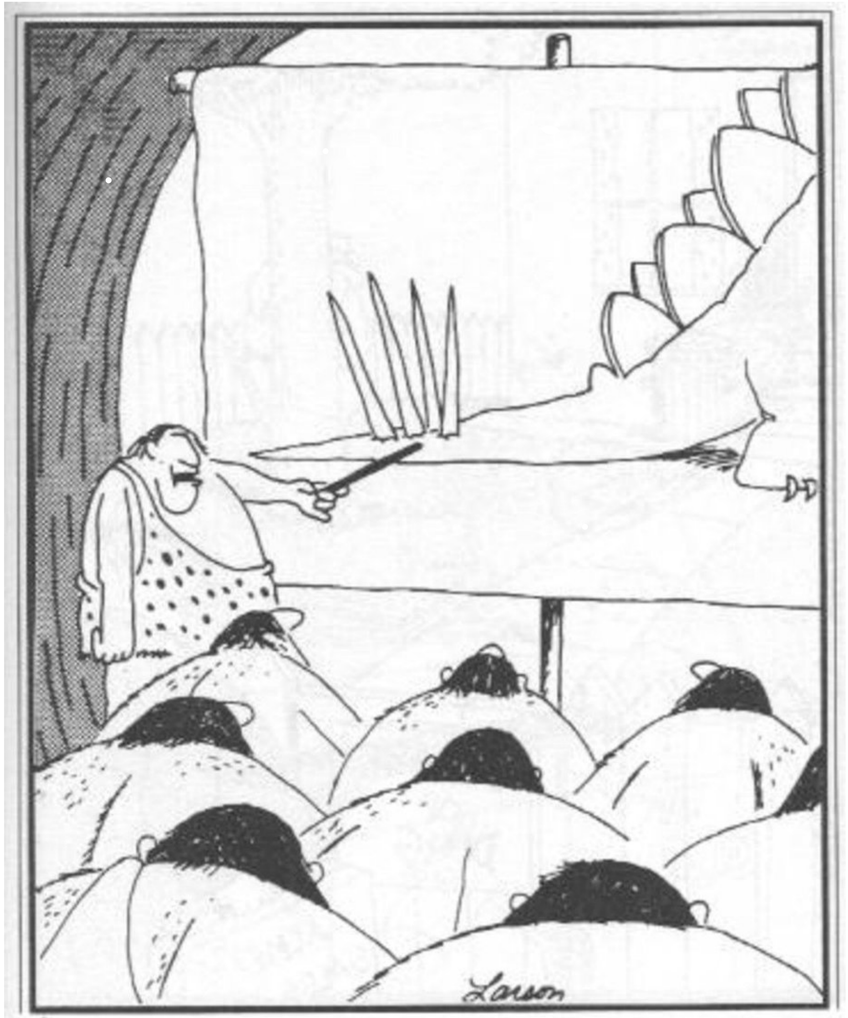
lots of different definitions





humans have used data forever

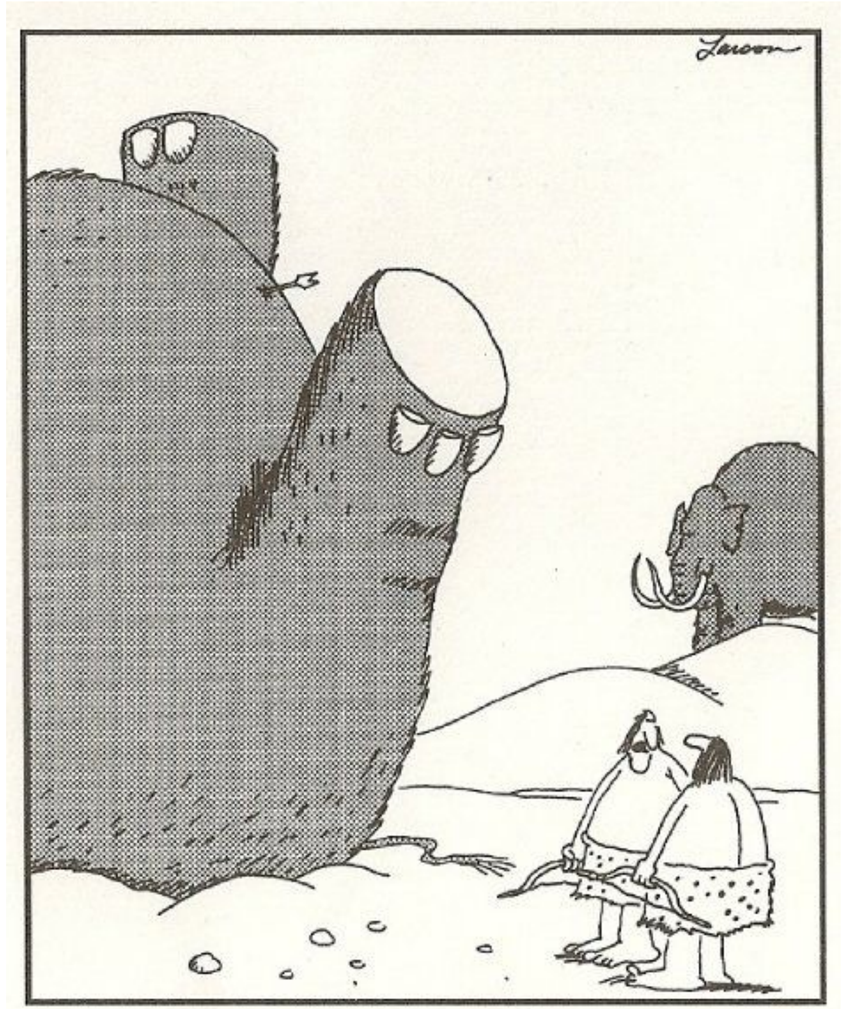
- Ever since Thag Simmons first thought, "Last time, we only sent two people to hunt the smilodon. Maybe this time we should send three?"



"Now this end is called the thagomizer . . . after the late Thag Simmons."

why do we use data?

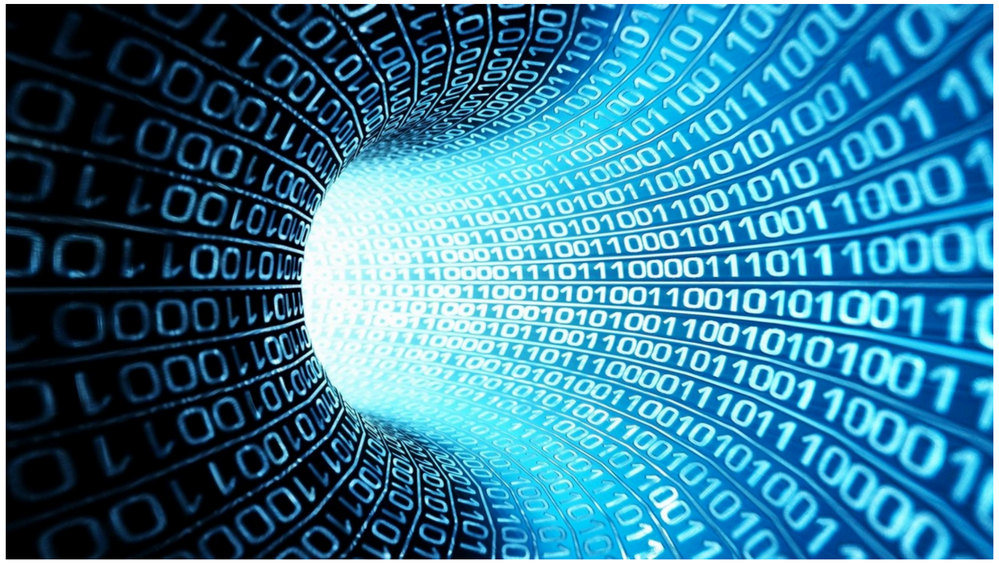
- Analyzing data helps us make decisions and take actions



"We should write that spot down."

what has changed?

- There's a lot more data
- Machines can also collect (and in turn use) it
- And we're trying to do more with it



a parable of purdue professors

Prof. Milind Kulkarni (ECE) builds systems
to make data analyses run faster

Prof. Bryan Pijanows
sound recordings fro
ecological change

Prof. Seungyoon Lee (Comm)
analyzes social media behavior
to understand how social networks
help people process information

Are they doing data science?

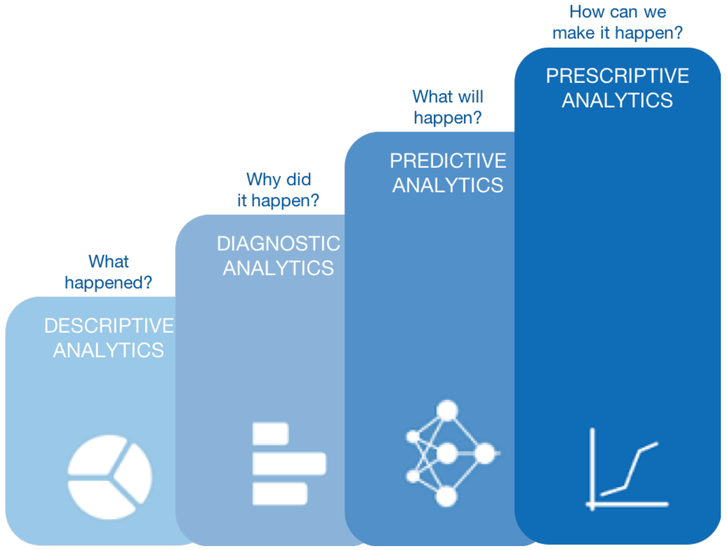
Neville (CS) builds
earning tools
pls and networks

Prof. Chris Brinton (ECE) develops
algorithms for modeling and
optimizing social and
communication networks from data

what is data science?

- Collecting data from a wide variety of sources and putting them into a consistent format?
- Making observations about patterns in data?
- Visualizing trends in data?
- Identifying similarities
- Making predictions abo in the future?
- Prescribing courses of ac to take based on forecasts?
- Developing new machine learning and data mining algorithms?
- Accelerating analysis algorithms?

Yes!



data science is a lot of things

making predictions from data

identifying patterns in data

building systems for data analysis

dealing with privacy concerns

visualizing data

collecting/organizing data

interpreting data

analyzing data

ethics

writing data analyses

data science is a lot of things

making predictions from data

identifying patterns in data

building systems for data analysis

dealing with privacy concerns

visualizing data

collecting/organizing data

interpreting data

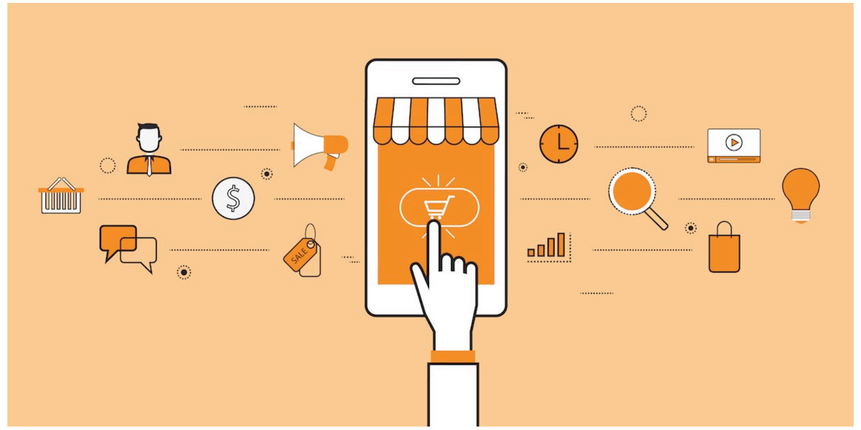
analyzing data

ethics

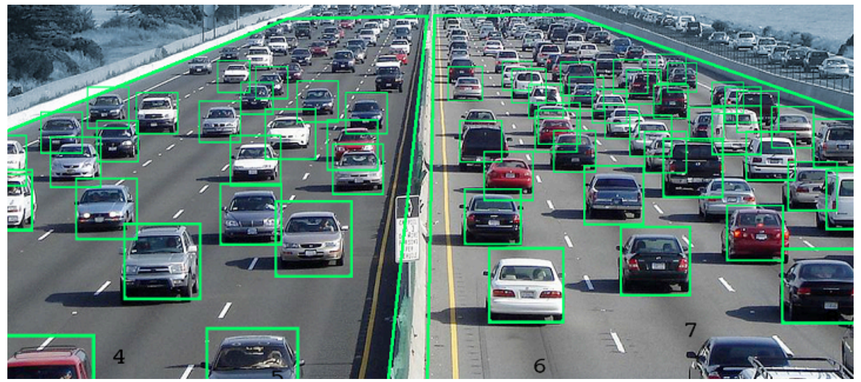
writing data analyses

what industries has it impacted?

- Hard to think of one that is *not* being impacted by data science!



- Medicine: Analytics from wearable trackers, studying disease patterns, …
- Retail: Analyzing consumer behavior, predicting customer satisfaction, …
- Transportation: Mapping customer journeys, predicting equipment failures, …
- Education: Tracking student engagement, personalizing learning content, …



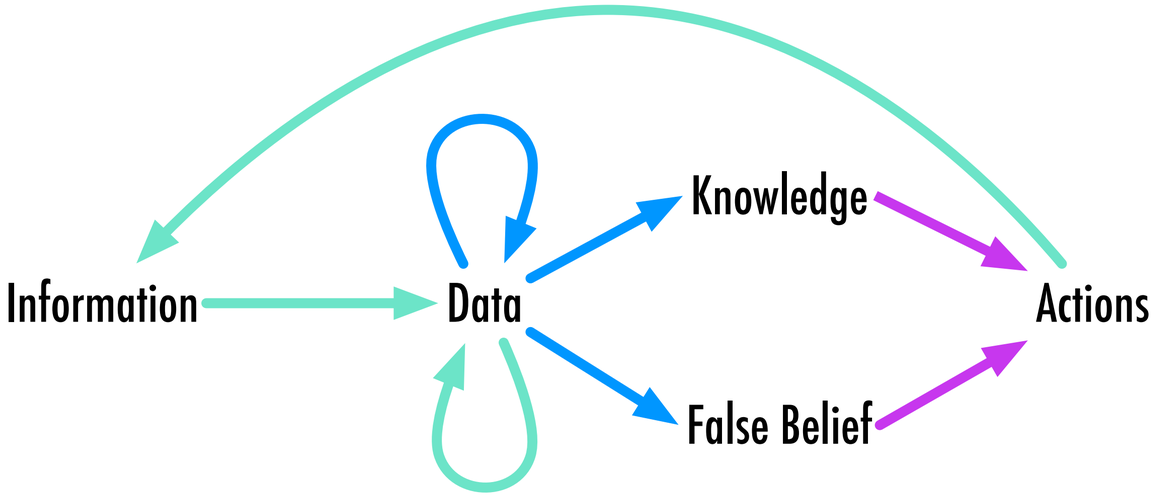
what about python?

- General purpose programming language, first appeared in the 90s
- Easily recognized by use of whitespace indentation rather than { } brackets to enhance readability
- Becoming the industry standard for data science (competing with R)
- Many useful, open-source libraries: numpy, pandas, matplotlib, pytorch
- And standard control functions (e.g., loops) from lower-level languages to help structure programs

```
59    # Build the TensorFlow graph.
60    g = tf.Graph()
61    with g.as_default():
62      # Build the model.
63      model = show_and_tell_model.ShowAndTellModel(
64          model_config, mode="train", train_inception=FLAGS.train_inception)
65      model.build()
66
67      # Set up the learning rate.
68      learning_rate_decay_fn = None
69      if FLAGS.train_inception:
70        learning_rate = tf.constant(training_config.train_inception_learning_rate)
71      else:
72        learning_rate = tf.constant(training_config.initial_learning_rate)
73        if training_config.learning_rate_decay_factor > 0:
74          num_batches_per_epoch = (training_config.num_examples_per_epoch /
75                                   model_config.batch_size)
76          decay_steps = int(num_batches_per_epoch *
77                            training_config.num_epochs_per_decay)
```

landscape

- This is an introductory programming course that emphasizes data science problems with some math
- Other data science courses in ECE:
  - ECE 30010 - Introduction to Machine Learning and Pattern Recognition
  - ECE 47300 - Introduction to Artificial Intelligence
  - ECE 59500 - Machine Learning I
- But data science is a Purdue-wide initiative!



syllabus break!