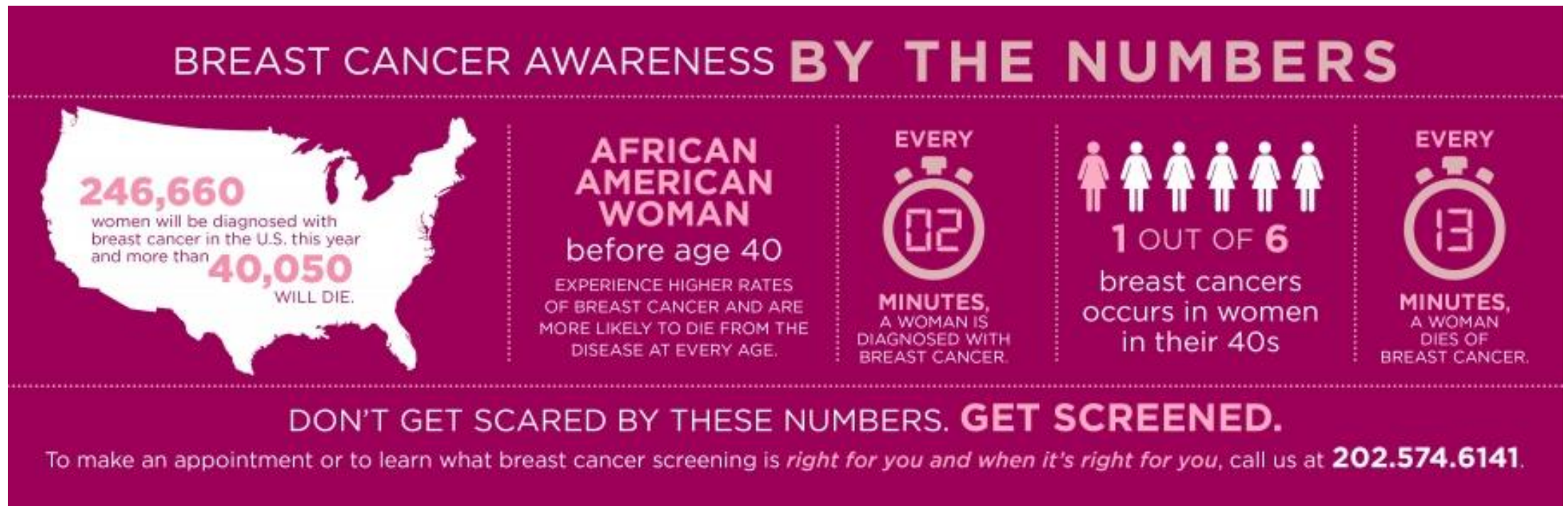CS+Social Good

CS 106S Winter 2018

# today's outline

1   breast cancer overview
2   intro to machine learning
3   cancer diagnosis assignment
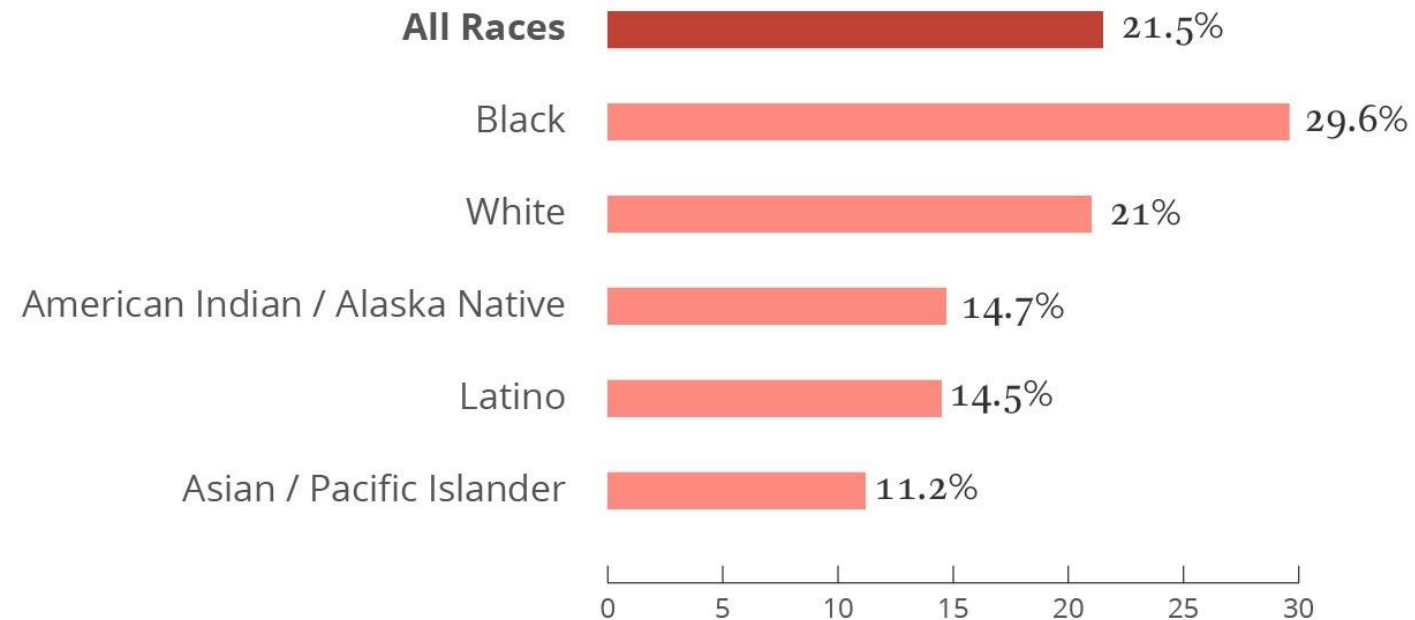
# breast cancer

# breast cancer

## U.S. Breast Cancer Deaths per 100,000 Women by Race/Ethnicity (2009-2013)

| | |
|---|---|
| **All Races** | 21.5% |
| Black | 29.6% |
| White | 21% |
| American Indian / Alaska Native | 14.7% |
| Latino | 14.5% |
| Asian / Pacific Islander | 11.2% |

# breast cancer

- error rates as high as **71%**
- mistakes are common depending on the doctor
- **31%** of all cases misdiagnosed
- **90,000** women misdiagnosed

# today's task

Given medical data about cell growths, can we accurately classify tumors as benign or malignant using machine learning?

# today's task **matters** because...

Classifying tumors allows:

- Doctors to diagnose cancer with **greater accuracy**
- Doctors and patients to make **more informed decisions**
- Medical costs to be reduced and diagnosing processes to become **more efficient**

# what is machine learning?

Machine learning is a "field of study that gives computers the ability to learn without being explicitly programmed."
-Arthur Samuel, 1959

# types of machine learning

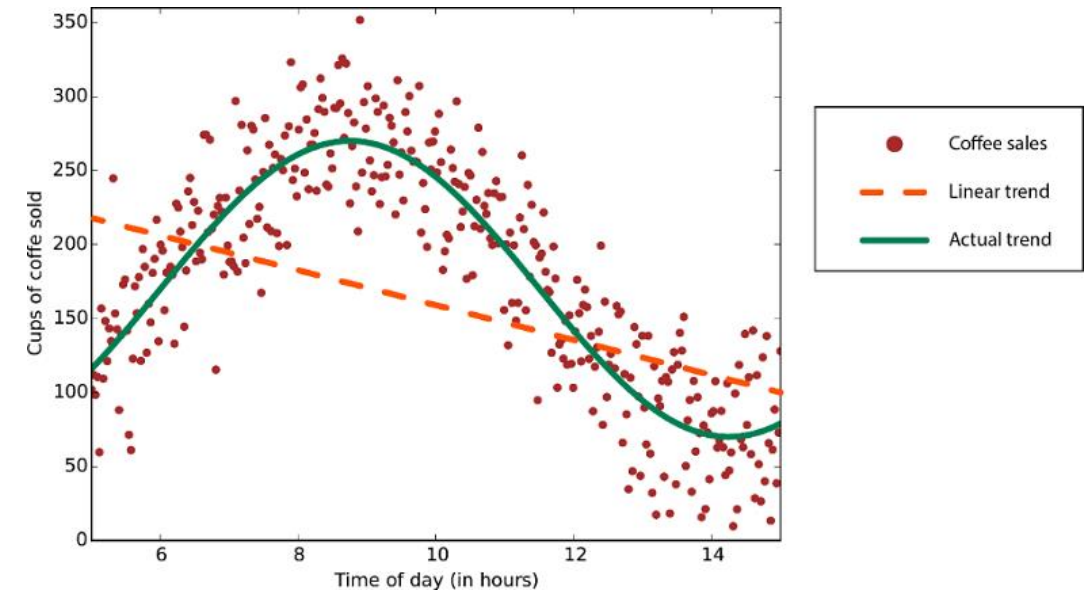Supervised

Unsupervised

# supervised machine learning
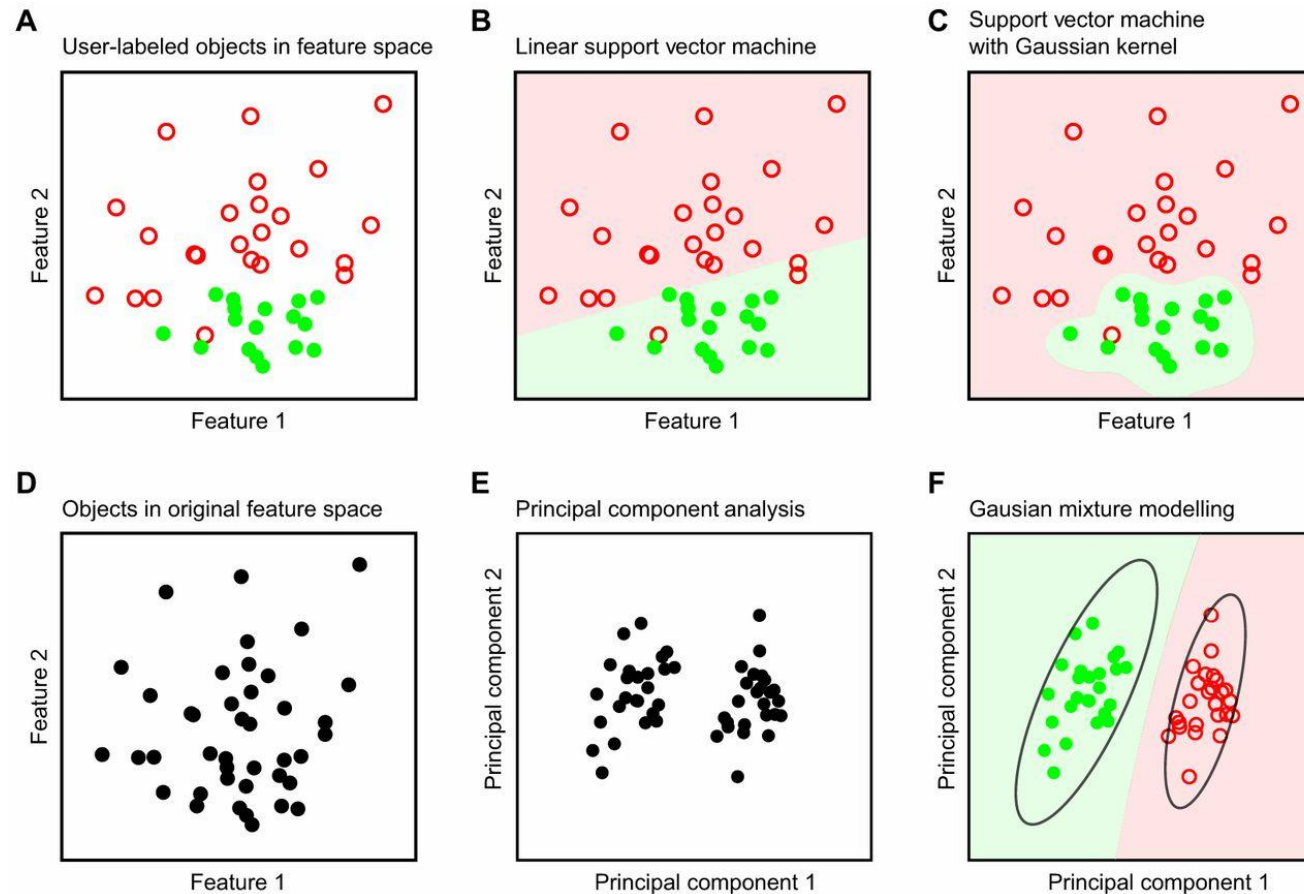
# unsupervised machine learning

# our data

- ~700 total data instances
- We are training on ~630, testing on ~70
- Each data instance has 11 values
  - Sample code number
  - Clump thickness (1-10)
  - Uniformity of Cell Size (1-10)
  - Uniformity of Cell Shape (1-10)
  - Marginal Adhesion (1-10)
  - **Label** (2=benign, 4=malignant)
  - Single Epithelial Cell Size (1-10)
  - Bare Nuclei (1-10)
  - Bland Chromatin (1-10)
  - Normal Nucleoli (1-10)
  - Mitosis (1-10)

# implementation steps

- Calculate distance between test instance and every train instance

- Pick the K train instances with the smallest distances

- Of these K train instances, see how many are classified as malignant and how many are classified as benign

- Pick whichever class appears more times as your answer

# let's get started!

**Build your own breast cancer classifier using JavaScript!**

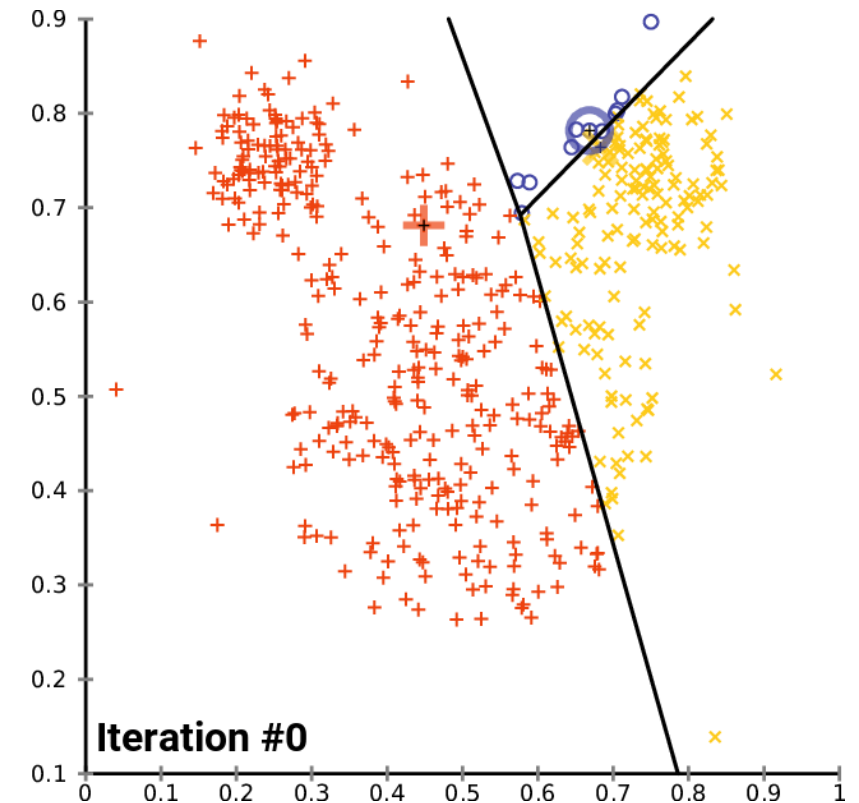Clone the code at `bit.ly/cs106srepo` and fill in `cancer-classify.js` located in `cancer-diagnosis/javascript`.

Run `python -m SimpleHTTPServer` in the above folder and examine the console output to test your solution.

# extra challenge

## Look up k-means and implement it with 2 clusters.

"k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster."

See you next time, friends!