

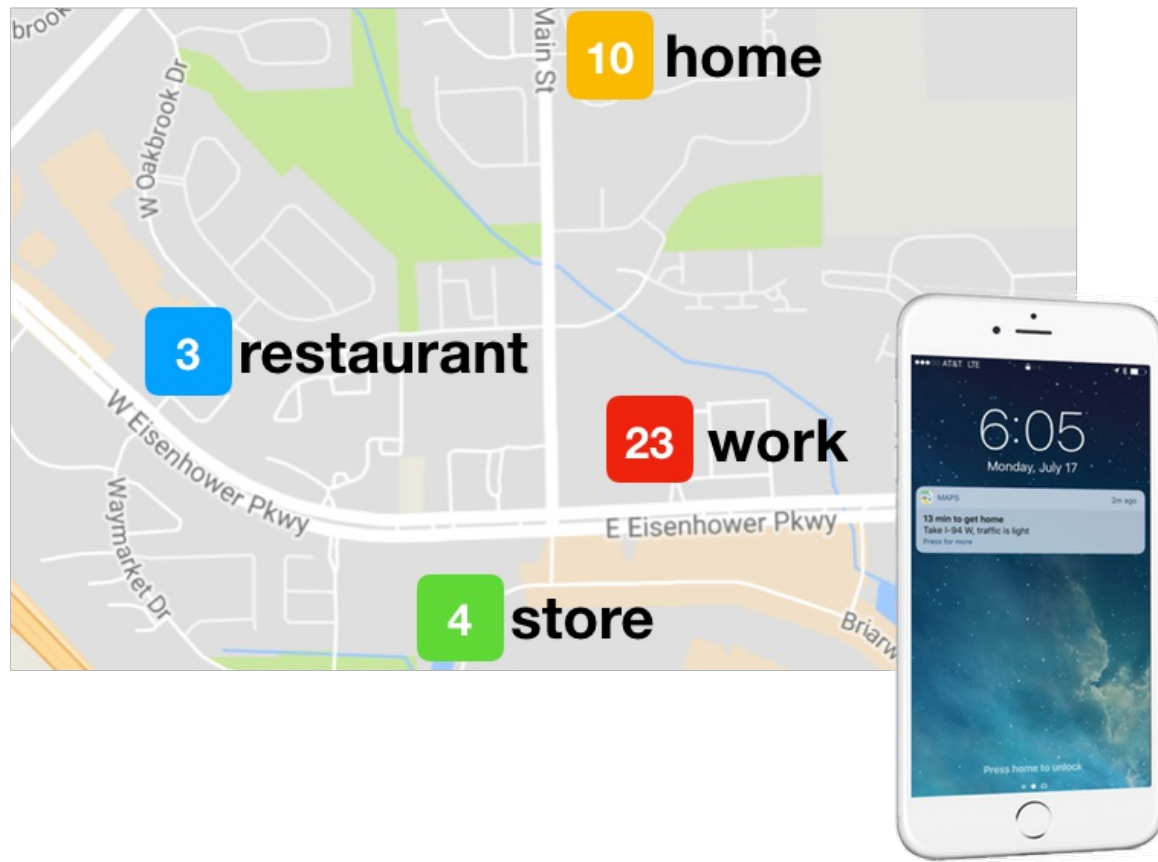


SUPERVISED LEARNING IN R: CLASSIFICATION

Understanding Bayesian methods

Brett Lantz
Instructor

Estimating probability



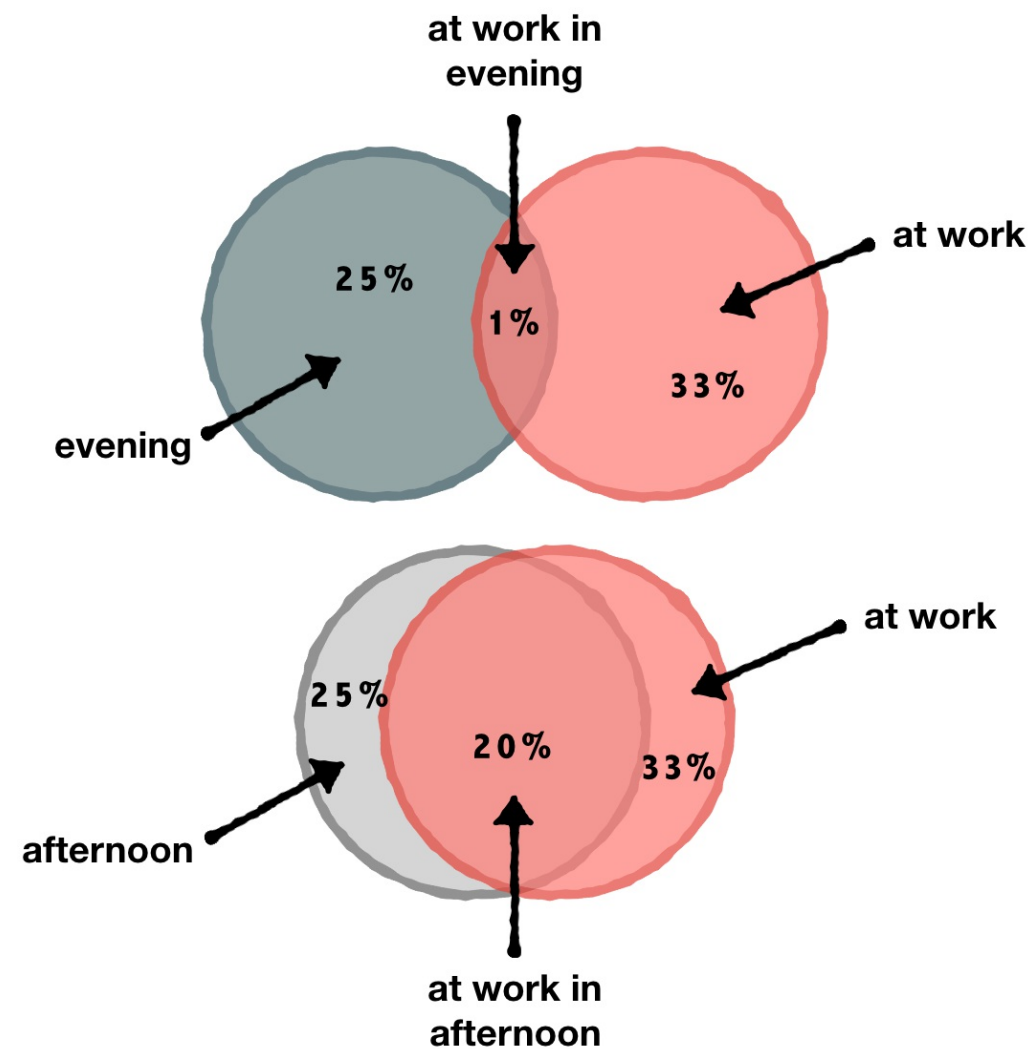
The **probability** of A is denoted $P(A)$

- $P(\text{work}) = 23 / 40 = 57.5\%$
- $P(\text{store}) = 4 / 40 = 10.0\%$

Joint probability and independent events

The **joint probability** of events A and B is denoted $P(A \text{ and } B)$

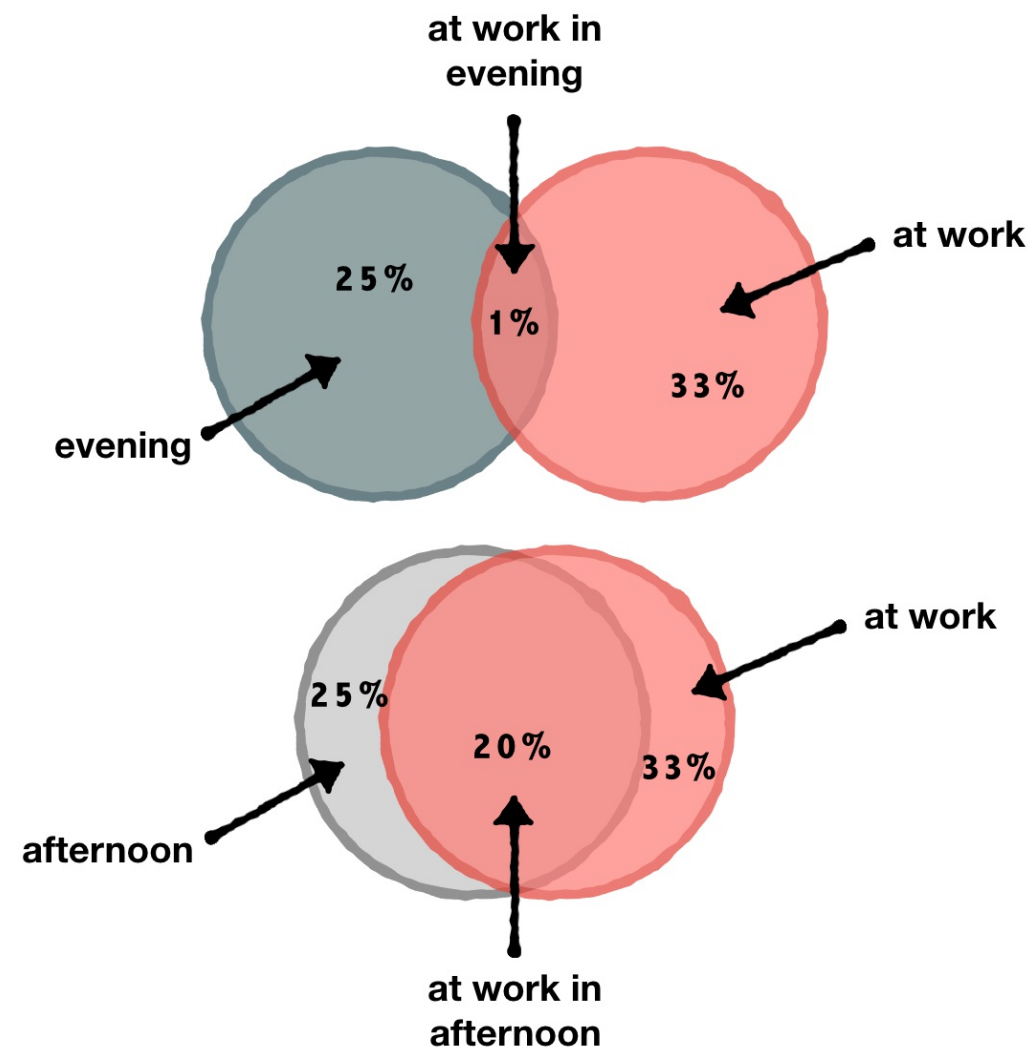
- $P(\text{work and evening}) = 1\%$
- $P(\text{work and afternoon}) = 20\%$



Conditional probability and dependent events

The **conditional probability** of events A and B is denoted $P(A | B)$

- $P(A | B) = P(A \text{ and } B) / P(B)$
- $P(\text{work} | \text{evening}) = 1 / 25 = 4\%$
- $P(\text{work} | \text{afternoon}) = 20 / 25 = 80\%$



Making predictions with Naive Bayes

```
# building a Naive Bayes model  
library(naivebayes)  
m <- naive_bayes(location ~ time_of_day, data = location_history)
```

```
# making predictions with Naive Bayes  
future_location <- predict(m, future_conditions)
```



SUPERVISED LEARNING IN R: CLASSIFICATION

Let's practice!

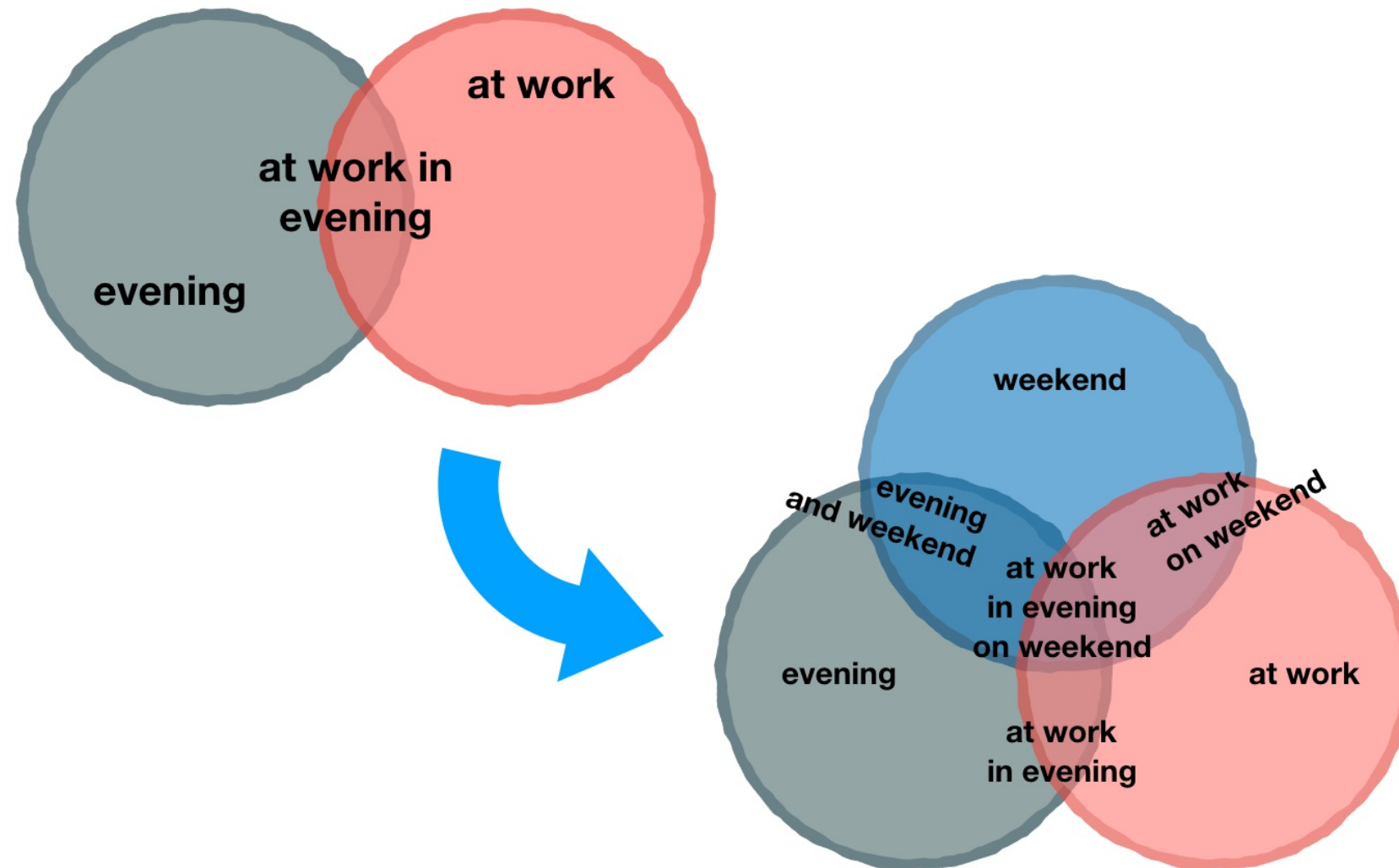


SUPERVISED LEARNING IN R: CLASSIFICATION

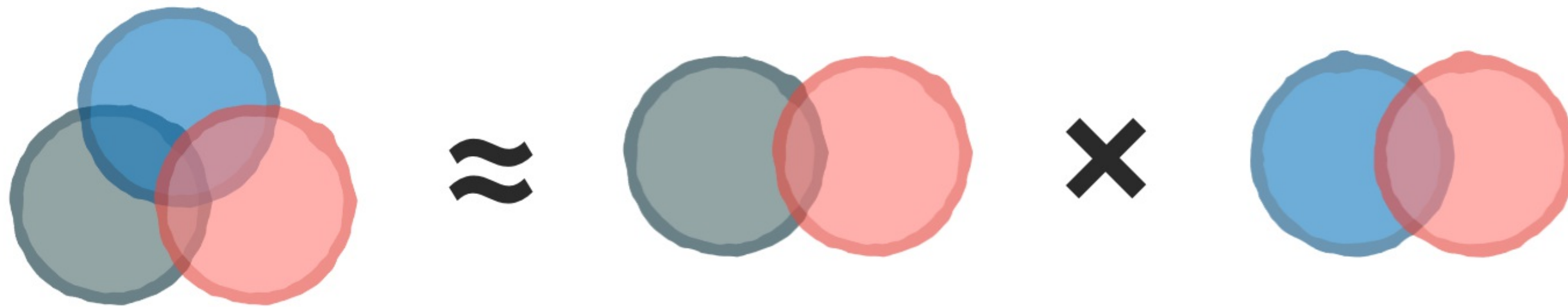
Understanding NB's "naivety"

Brett Lantz
Instructor

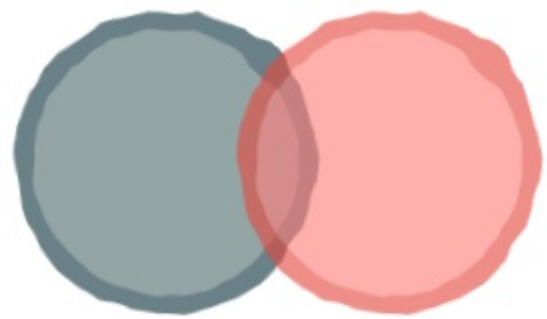
The challenge of multiple predictors



A "naive" simplification

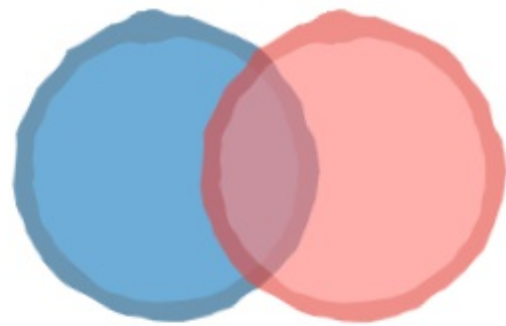


An "infrequent" problem



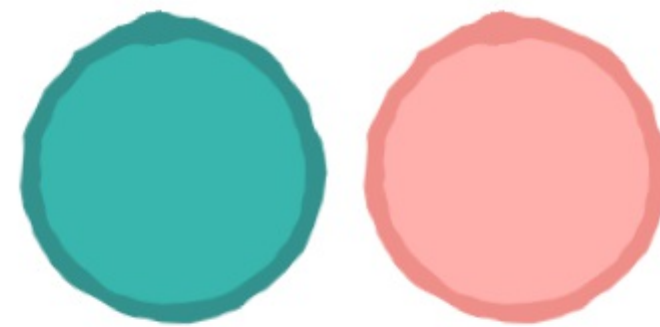
$$P(A \cap B) = 0.10$$

×



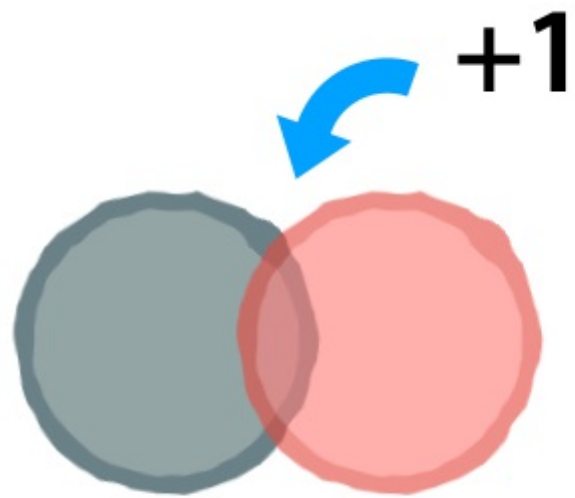
$$P(A \cap B) = 0.30$$

×



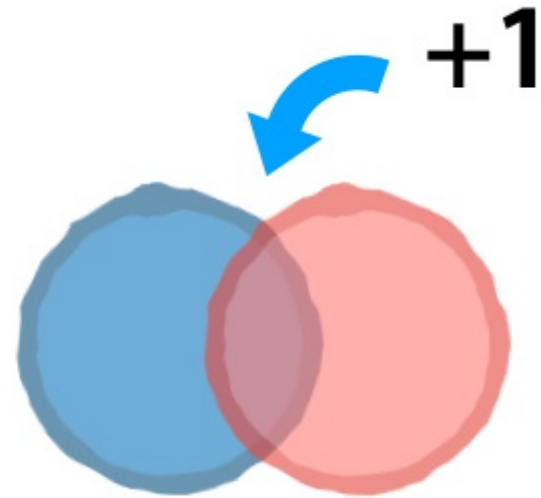
$$P(A \cap B) = 0.00$$

The Laplace correction



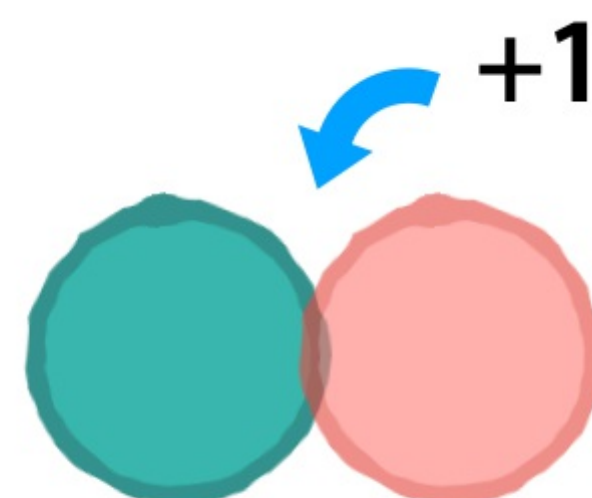
$$P(A \cap B) = 0.11$$

×



$$P(A \cap B) = 0.30$$

×



$$P(A \cap B) = 0.01$$



SUPERVISED LEARNING IN R: CLASSIFICATION

Let's practice!



SUPERVISED LEARNING IN R: CLASSIFICATION

Applying Naive Bayes to other problems

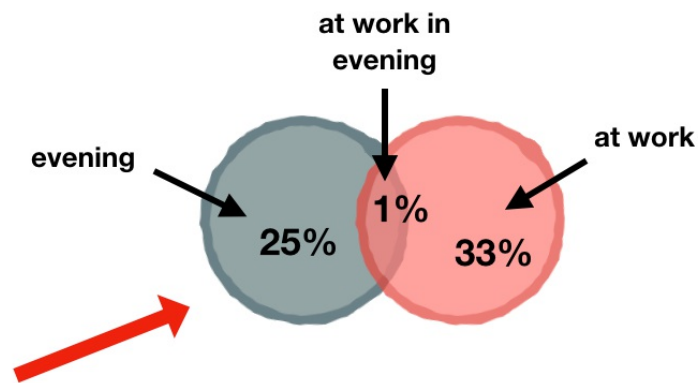
Brett Lantz
Instructor

How Naive Bayes uses data

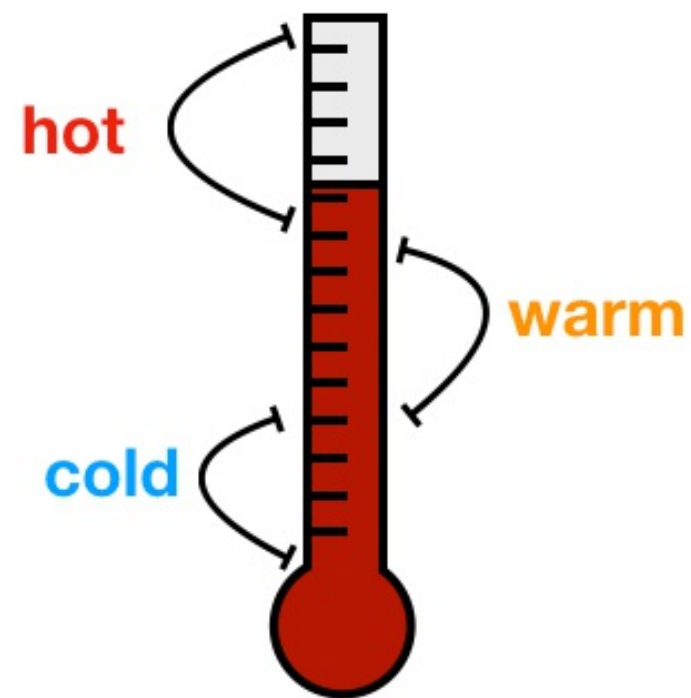
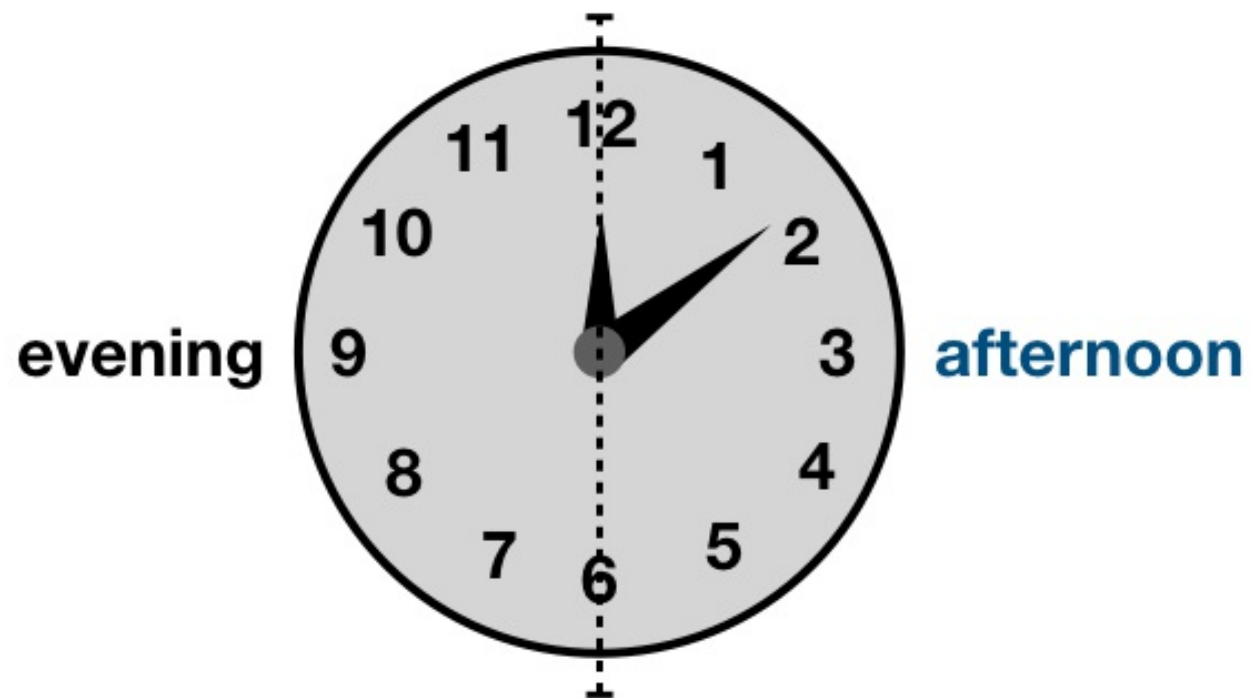
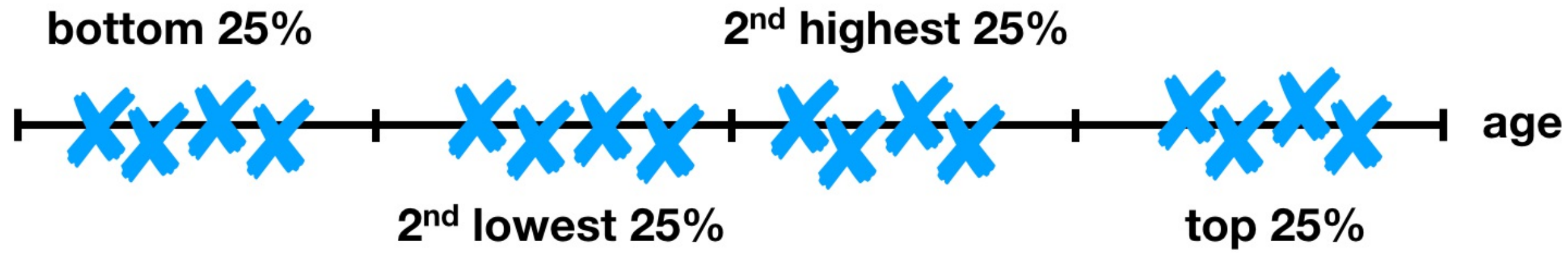
	month	day	weekday	daytype	hour	hourtype	location
1	1	4	wednesday	weekday	0	night	home
2	1	4	wednesday	weekday	1	night	home
3	1	4	wednesday	weekday	2	night	home
4	1	4	wednesday	weekday	3	night	home
5	1	4	wednesday	weekday	4	night	home
6	1	4	wednesday	weekday	5	night	home
7	1	4	wednesday	weekday	6	morning	home
8	1	4	wednesday	weekday	7	morning	home
9	1	4	wednesday	weekday	8	morning	home
10	1	4	wednesday	weekday	9	morning	office
11	1	4	wednesday	weekday	10	morning	office
12	1	4	wednesday	weekday	11	morning	office
13	1	4	wednesday	weekday	12	afternoon	office



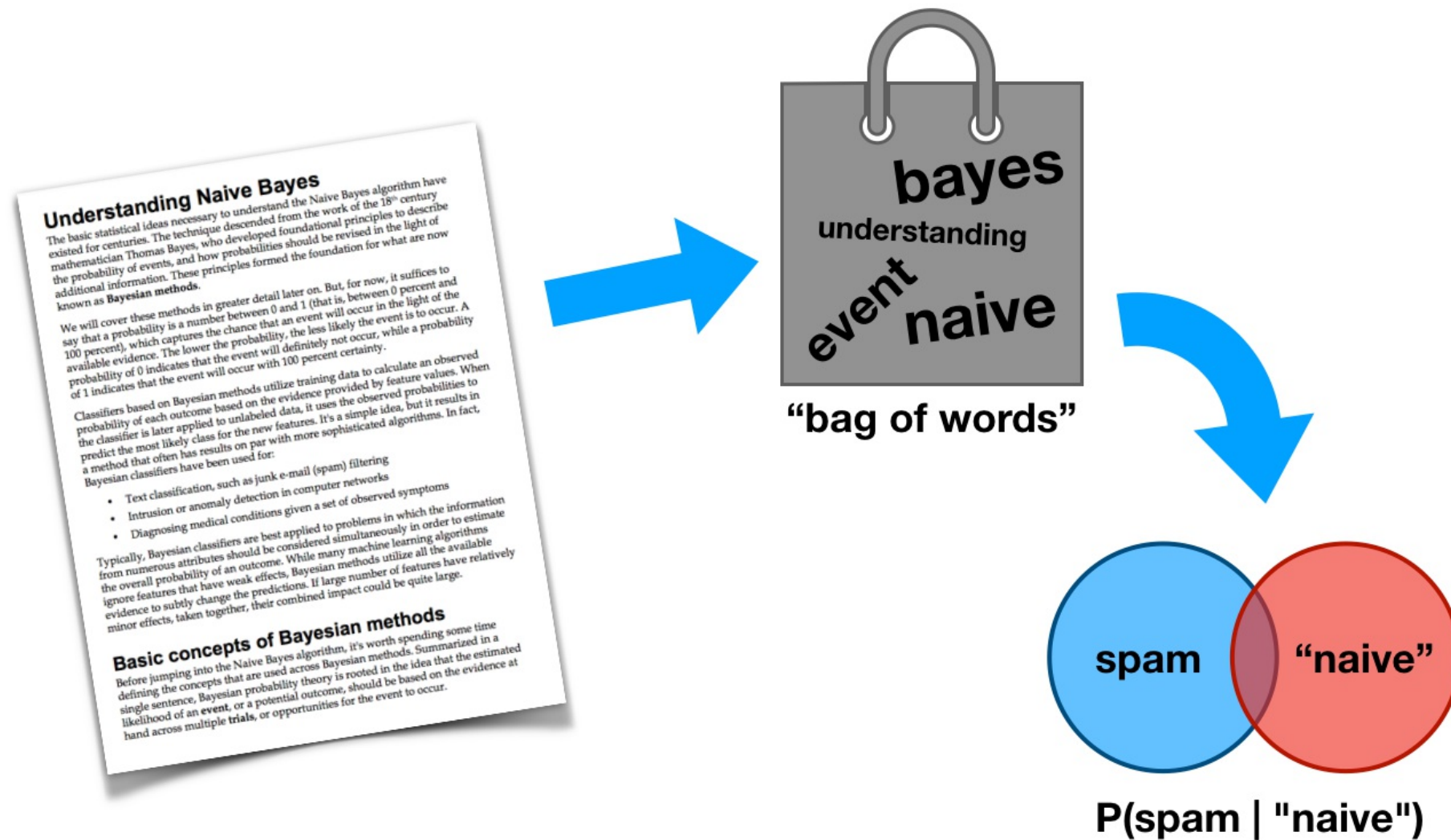
	not evening	evening	
at work	705	22	
not at work	933	524	



Binning numeric data for Naive Bayes



Preparing text data for Naive Bayes





SUPERVISED LEARNING IN R: CLASSIFICATION

Let's practice!