

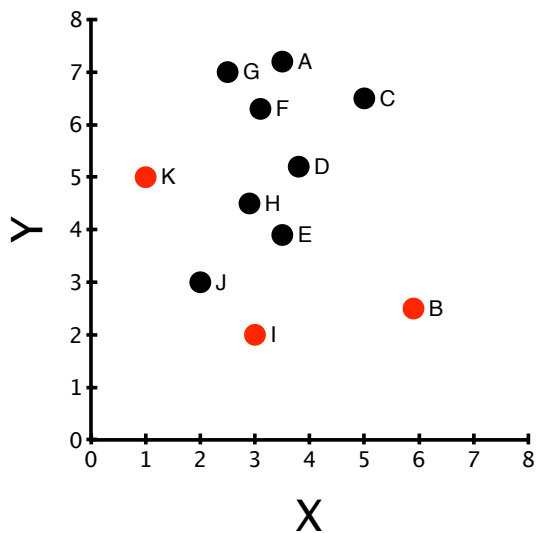
# Tools & Models for Data Science

## Outliers Handout

```

init min-priority queue  $O$ 
for  $x_1 \in X$ :
  init max-priority queue  $Q$ 
  for  $x_2 \neq x_1 \in X$ :
    insert  $dist(x_1, x_2)$  into  $Q$ 
    if  $|Q| > k$ 
      remove max from  $Q$ 
  insert  $x_1$  into  $O$  with key  $\max(Q)$ 
  if  $|O| > m$ 
    remove point with min key from  $O$ 
return  $O$ 

```



Point	2NN distance
A	1.02
B	2.94
C	1.77
D	1.30
E	1.33
F	0.98
G	1.02
H	1.14
I	1.96
J	1.75
K	2.24

```

O: { }
Pick a point: A
Q: { }
    Pick a point: B
    Q: {5.28}
    Q is not full

    Pick a point: C
    Q: {1.66, 5.28}
    Q is not full

    Pick a point: D
    Q: {1.66, 2.02, 5.28}
    Q is full
    Q: {1.66, 2.02}

    Pick a point: E
    Q: {1.66, 2.02, 3.30}
    Q is full
    Q: {1.66, 2.02}
    ...
    Q: {0.98, 1.02}

O: {(1.02, A)}
O is not full

Next point: B
Q: { }
    Pick a point: A
    Q: {5.28}
    Q is not full

    Pick a point: C
    Q: {4.1, 5.28}
    Q is not full

    Pick a point: D
    Q: {1.77, 4.1, 5.28}
    Q is full
    Q: {1.77, 4.1}
    ...

    Q: {2.78, 2.94}
O: {(1.02, A), (2.94, B)}

Finally, get
O: {(2.94, B), (2.24, K), (1.96, I)}

```