

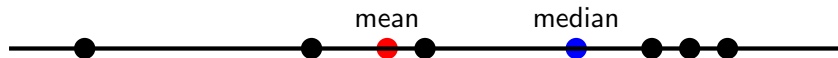
Dynamically Computing Depth in Data Sets

Sydney Leither (Baldwin Wallace University),
Sunaina Butler, Ben Mikek (Grinnell College)

COURAGE @ Clemson University

MATH REU Conference @ Clemson University, July 2020

The 'Center' of a Data Set



The median is more robust to outliers than the mean

Data Depth

Definition

Depth of data points: proximity to center

Definition

Deepest point: generalization of the median

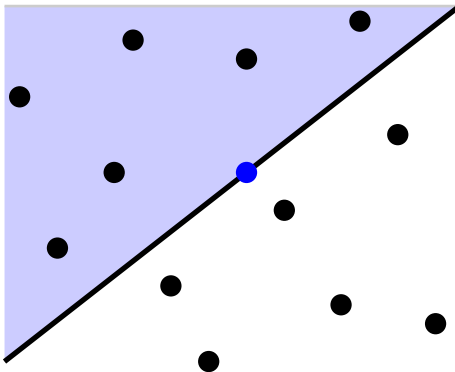
- generalize the median to higher dimensions
- cannot be calculated as easily

Half-Space Depth

Definition

Half-space: one side of a line drawn through a point

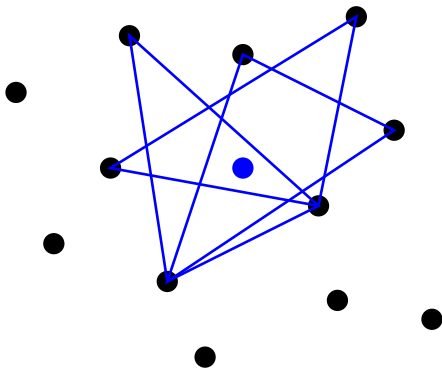
Where can we draw any line and evenly split the data set?



Measure of depth: minimum number of points in a half-space

Simplicial Depth

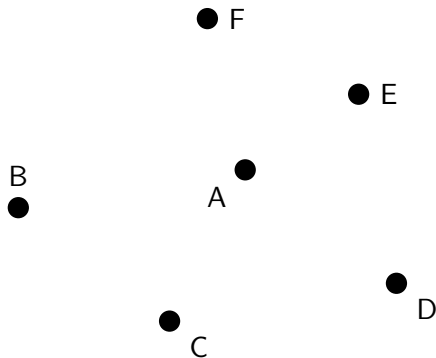
Which point is in the interior of the most triangles?



Measure of depth: number of triangles

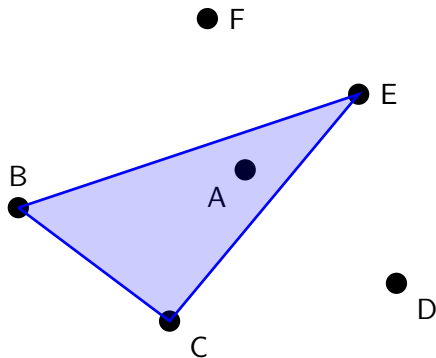
Guess the Depth!

- Point A: deepest point
- How many extra triangles does it get?
- Note that:
 - ▶ $\binom{6}{3} = 20$
 - ▶ $\binom{5}{2} = 10$



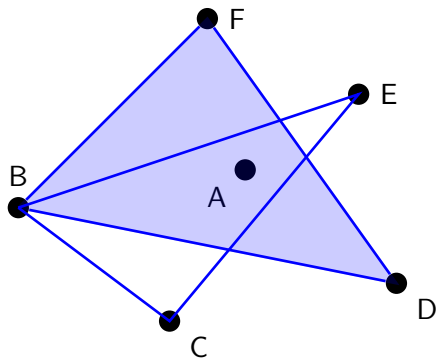
Guess the Depth!

- Point A: deepest point
- How many extra triangles does it get?
- Note that:
 - ▶ $\binom{6}{3} = 20$
 - ▶ $\binom{5}{2} = 10$



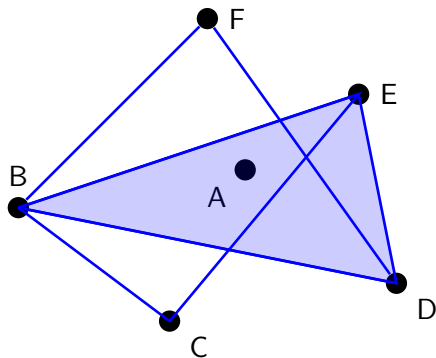
Guess the Depth!

- Point A: deepest point
- How many extra triangles does it get?
- Note that:
 - ▶ $\binom{6}{3} = 20$
 - ▶ $\binom{5}{2} = 10$



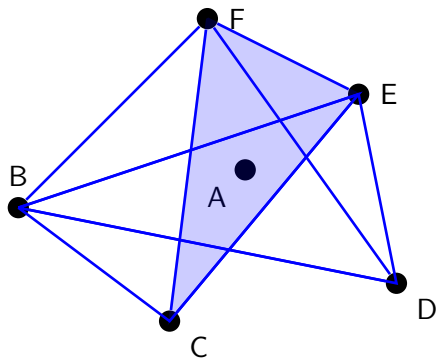
Guess the Depth!

- Point A: deepest point
- How many extra triangles does it get?
- Note that:
 - ▶ $\binom{6}{3} = 20$
 - ▶ $\binom{5}{2} = 10$



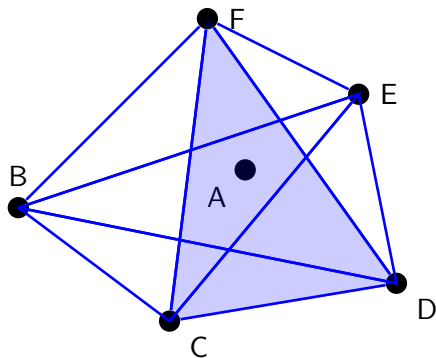
Guess the Depth!

- Point A: deepest point
- How many extra triangles does it get?
- Note that:
 - ▶ $\binom{6}{3} = 20$
 - ▶ $\binom{5}{2} = 10$



Guess the Depth!

- Point A: deepest point
- How many extra triangles does it get?
- Note that:
 - ▶ $\binom{6}{3} = 20$
 - ▶ $\binom{5}{2} = 10$

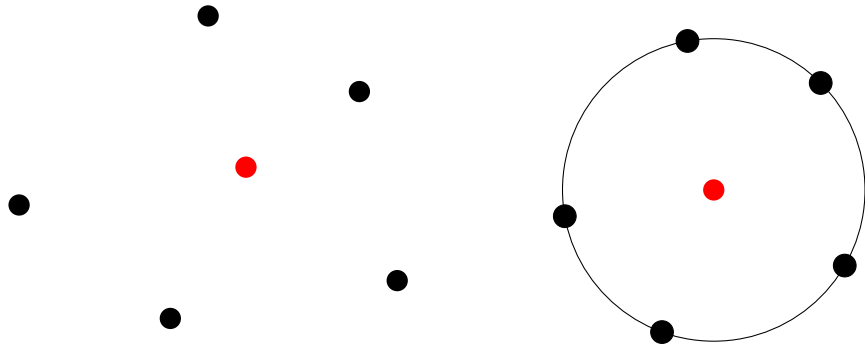


Static Simplicial Depth

- Baseline for dynamic algorithm
- Goal: Find the simplicial depth of a point within a dataset
- Idea: Find the total number of triangles and subtract the triangles that do not contain the query point

Static Simplicial Depth Algorithm

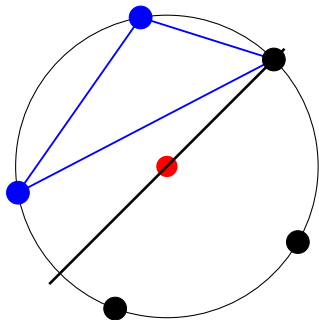
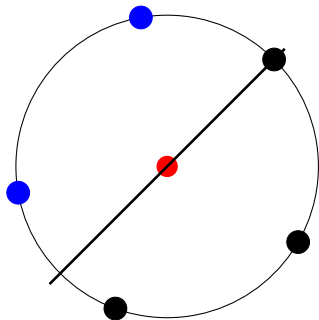
- Project points onto circle centered at query point
- Sort points radially
- Count points (n)



$n = 5$

Static Simplicial Depth Algorithm: Half-Space

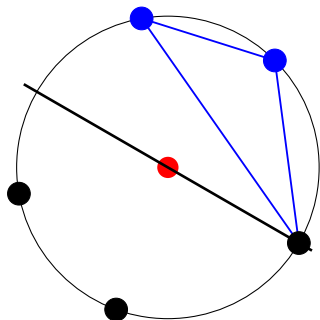
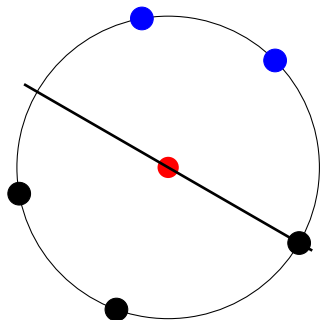
- Until back at start:
 - ▶ Store number of points strictly left of the bounding line (h_i)
 - ▶ Rotate bounding line clockwise until a new point is hit



$$h_1 = 2$$

Static Simplicial Depth Algorithm: Half-Space

- Until back at start:
 - ▶ Store number of points strictly left of the bounding line (h_i)
 - ▶ Rotate bounding line clockwise until a new point is hit



$$h_1 = 2$$

$$h_2 = 2$$

Static Simplicial Depth

Plug into formula $\binom{n}{3} - \sum_{i=1}^n \binom{h_i}{2}$

$$n = 5$$

$$h_1 = 2$$

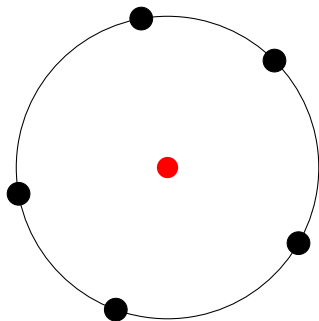
$$h_2 = 2$$

$$h_3 = 2$$

$$h_4 = 2$$

$$h_5 = 2$$

$$\text{Depth} = \binom{5}{3} - \left(\binom{2}{2} + \binom{2}{2} + \binom{2}{2} + \binom{2}{2} + \binom{2}{2} \right) = 5$$



A Dynamic Algorithm for Simplicial Depth

Problem

Fix the query point θ and add a point x .

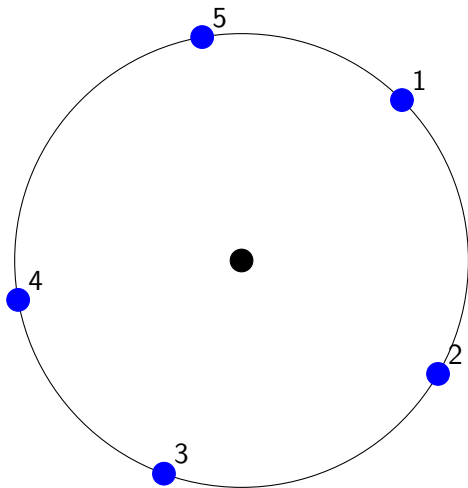
What is the new depth of θ ?

- 1 **Increment the number of points for half-spaces that contain x**
- 2 **Compute the number of points in x 's half-space**
- 3 **Add to the data set**

Dynamic Algorithm: Part 1

- 1 Increment the number of points for half-spaces that contain x

data points

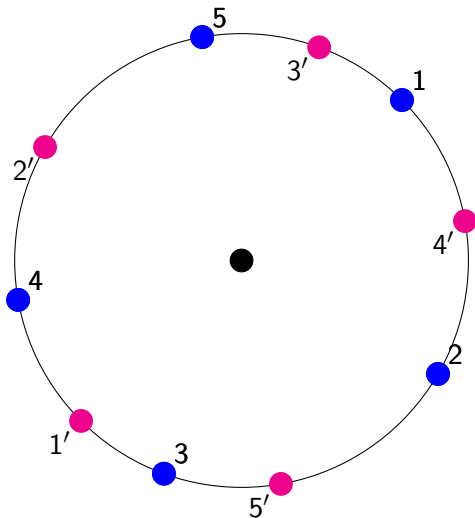


Dynamic Algorithm: Part 1

- 1 Increment the number of points for half-spaces that contain x

data points

ghost points



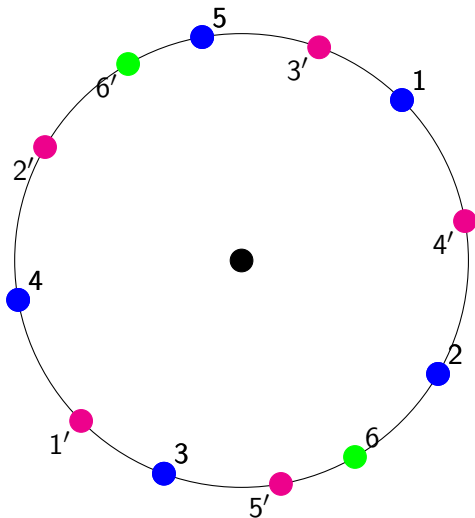
Dynamic Algorithm: Part 1

- 1 Increment the number of points for half-spaces that contain x

data points

ghost points

new point



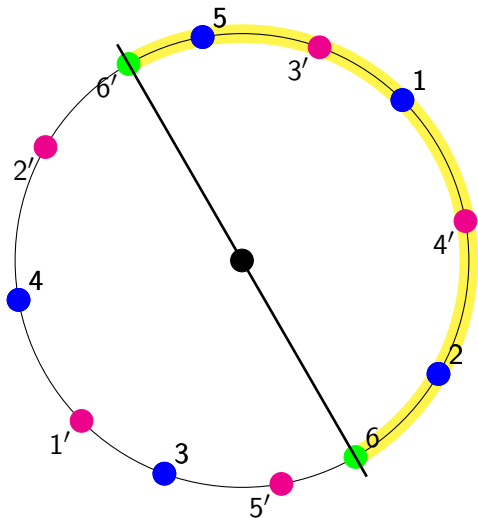
Dynamic Algorithm: Part 1

- 1 Increment the number of points for half-spaces that contain x

data points

ghost points

new point



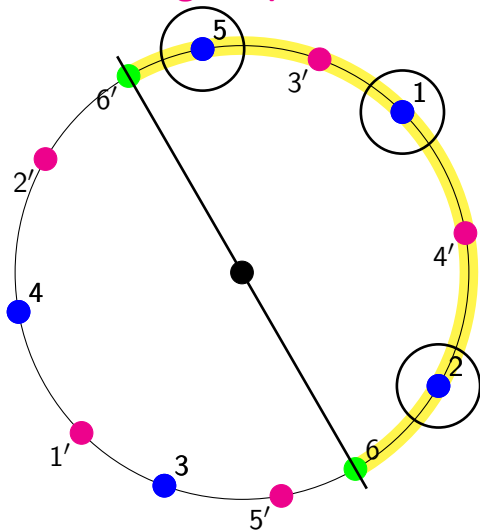
Dynamic Algorithm: Part 1

- 1 Increment the number of points for half-spaces that contain x

data points

ghost points

new point



Dynamic Algorithm: Part 1

- 1 **Increment the number of points for half-spaces that contain x**

Lemma

For some set of values $\{k_i\}_{i=1}^n$ with $k_i \in \mathbb{N}$,

$$\sum_{i=1}^n \binom{k_i + 1}{2} = \sum_{i=1}^n \binom{k_i}{2} + \sum_{i=1}^n k_i$$

Dynamic Algorithm: Part 1

- 1 **Increment the number of points for half-spaces that contain x**

Lemma

For some set of values $\{k_i\}_{i=1}^n$ with $k_i \in \mathbb{N}$,

$$\sum_{i=1}^n \binom{k_i + 1}{2} = \sum_{i=1}^n \binom{k_i}{2} + \sum_{i=1}^n k_i$$

Point: 5 1 2

Before: $\binom{2}{2} + \binom{2}{2} + \binom{2}{2} = 3$

Dynamic Algorithm: Part 1

- 1 Increment the number of points for half-spaces that contain x

Lemma

For some set of values $\{k_i\}_{i=1}^n$ with $k_i \in \mathbb{N}$,

$$\sum_{i=1}^n \binom{k_i + 1}{2} = \sum_{i=1}^n \binom{k_i}{2} + \sum_{i=1}^n k_i$$

Point:	5	1	2	
Before:	$\binom{2}{2}$	$\binom{2}{2}$	$\binom{2}{2}$	= 3
After:	$\binom{3}{2}$	$\binom{3}{2}$	$\binom{3}{2}$	$\underbrace{= 9}_{3+(2+2+2)=9}$

A Dynamic Algorithm for Simplicial Depth

Problem

Fix the query point θ and add a point x .

What is the new depth of θ ?

- 1 Increment the number of points for half-spaces that contain x
- 2 **Compute the number of points in x 's half-space**
- 3 Add to the data set

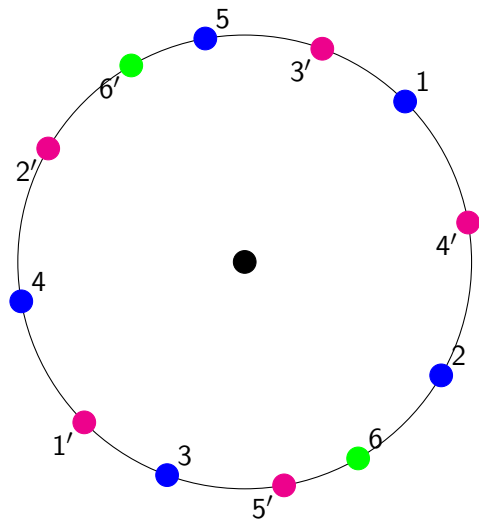
Dynamic Algorithm: Part 2

- 2 Compute the number of points in x 's half-space

data points

ghost points

new point



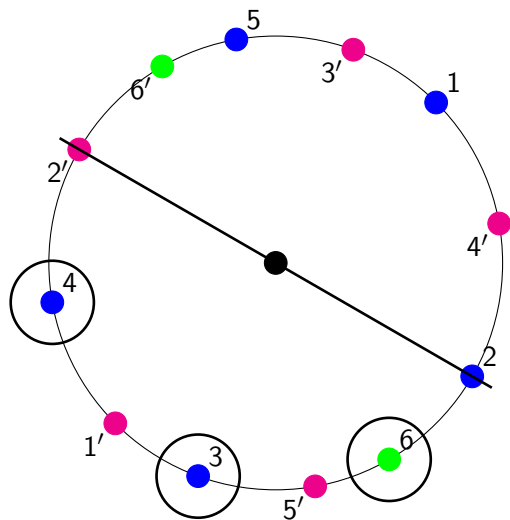
Dynamic Algorithm: Part 2

- 2 Compute the number of points in x 's half-space

data points

ghost points

new point



Initial: 3 points

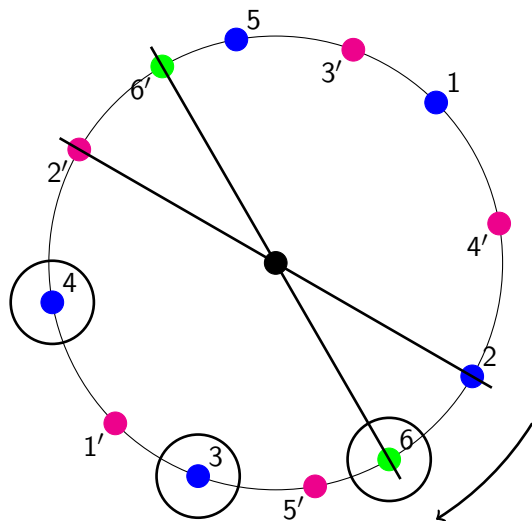
Dynamic Algorithm: Part 2

② Compute the number of points in x 's half-space

data points

ghost points

new point



Initial: 3 points

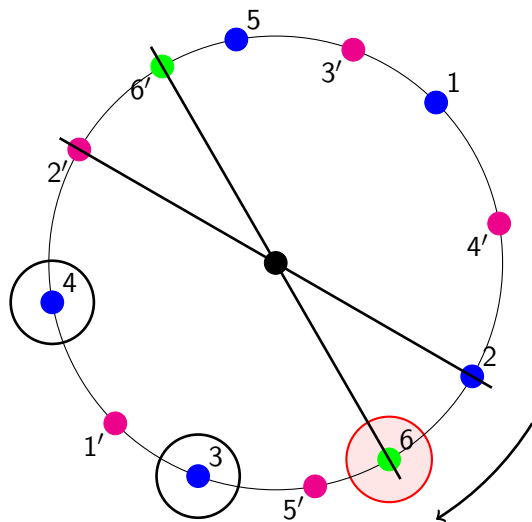
Dynamic Algorithm: Part 2

② Compute the number of points in x 's half-space

data points

ghost points

new point



Initial: 3 points
Remove point 6

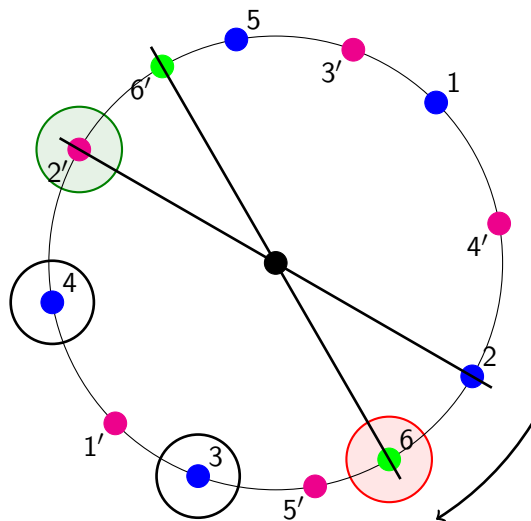
Dynamic Algorithm: Part 2

② Compute the number of points in x 's half-space

data points

ghost points

new point



Initial: 3 points

Remove point 6

Add point 2'

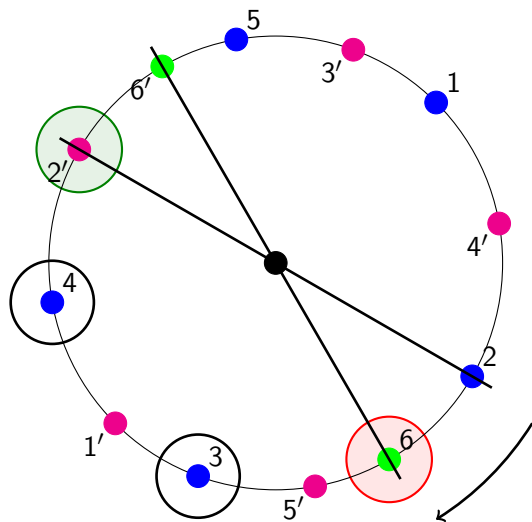
Dynamic Algorithm: Part 2

② Compute the number of points in x 's half-space

data points

ghost points

new point



Initial: 3 points

Remove point 6

Add point 2'

Final: 2 points

A Dynamic Algorithm for Simplicial Depth

Problem

Fix the query point θ and add a point x .

What is the new depth of θ ?

- 1 **Increment the number of points for half-spaces that contain x**
- 2 **Compute the number of points in x 's half-spaces**
- 3 **Add to the data set**

Dynamic Algorithm: Part 3

③ Add x to the data set

Counting triangles that **do not** contain θ :

	Before 6		Contributed by 6		After 6		
Before:	3	+	0	+	2	=	5

Dynamic Algorithm: Part 3

3 Add x to the data set

Counting triangles that **do not** contain θ :

	Before 6		Contributed by 6		After 6		
Before:	3	+	0	+	2	=	5
After:	9	+	$\binom{2}{2} = 1$	+	2	=	12

Dynamic Algorithm: Part 3

3 Add x to the data set

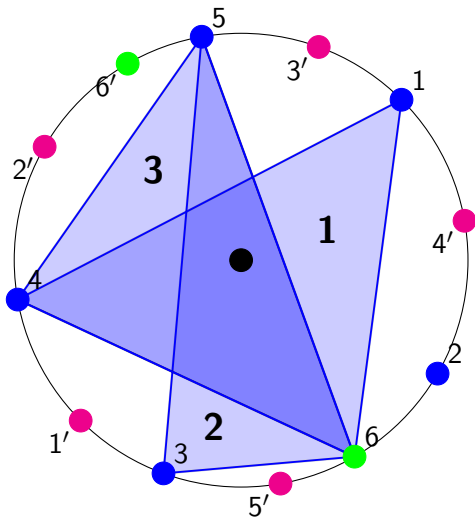
Counting triangles that **do not** contain θ :

	Before 6		Contributed by 6		After 6		
Before:	3	+	0	+	2	=	5
After:	9	+	$\binom{2}{2} = 1$	+	2	=	12

$$\binom{5}{3} - 5 = 5 \quad \longrightarrow \quad \binom{6}{3} - 12 = 8$$

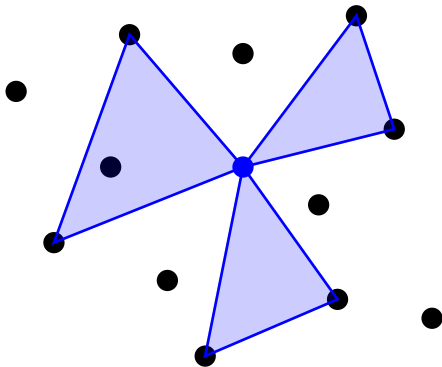
Reasonableness Check

$$\binom{5}{3} - 5 = 5 \quad \longrightarrow \quad \binom{6}{3} - 12 = 8$$



Future Questions

- A dynamic algorithm for Oja depth and other depth measures



- How are the different depth measures related?

Thank you!