

ANALYSE FACTORIELLE DES CORRESPONDANCES

pour l'analyse des variable qualitatives

encadrée par : Pr AGHRICH

par : ZAHID ACHRAF & MADIH MOHAMED NIZAR





C'est quoi l'AFC



definition



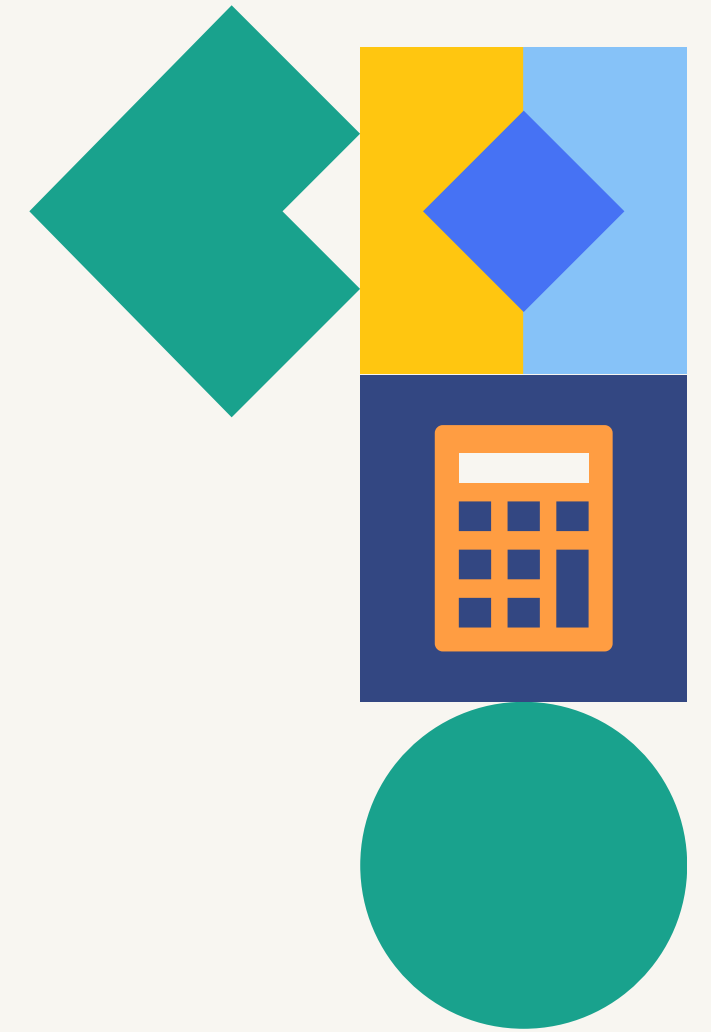
etapes

Answer Key

AFC?

L'analyse factorielle des correspondances (AFC) est une méthode qui étudie les liens entre deux variables qualitatives. Elle transforme un tableau de données en une représentation graphique où les catégories proches sont liées. L'AFC réduit l'information tout en mettant en valeur les correspondances importantes. Contrairement à l'ACP, elle travaille sur des variables qualitatives et offre un espace commun pour individus et modalités.

C and D have the same mean of 13.



ACP vs AFC

différence entre ACP et AFC

Critère	ACP	AFC
type données	Quantitatives	Qualitatives
donnée utilisée	Données numériques	Tableaux de contingence
Objectif	Résumer la variance des données	Étudier les correspondances entre catégories
PRESENTATION	Seulement individus	Individus et modalités



ACP? AFC?



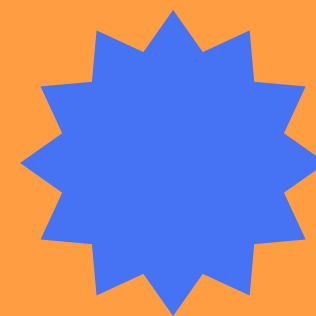


usage



AFC

données qualitatives, liens
entre catégories.



ACP

données quantitatives,
réduction de la dimension
en maximisant la variance.

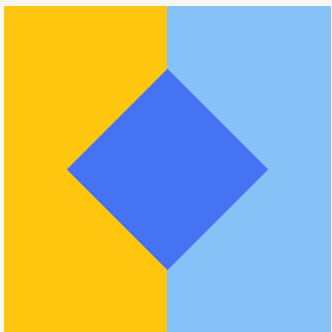




Les étapes de l'AFC



Étape 1 : Construction du tableau de données



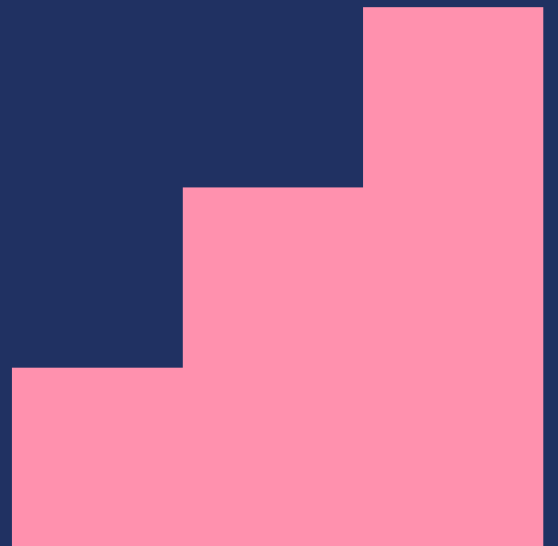
L'AFC nécessite un tableau de données croisées appelé "tableau de contingence". Ce tableau doit être construit à partir des données brutes, en croisant les valeurs de chaque variable.

sexe	travaile		Color	ing	doc	direc	dev	
femme	doc		homme	101	29	201	550	...
homme	doc		femme	82	130	92	478	
homme	dev							
femme	ing							
	⋮							

Étape 2 : Calcul de la distance de khi-deux

La distance de khi-deux est une mesure de la différence entre les observations et les fréquences attendues sous l'hypothèse d'indépendance entre les variables. Elle est utilisée pour mesurer la similitude entre deux variables. La formule de la distance de khi-deux est la suivante :

où $n_{i.}$ est la somme des effectifs de la ligne i , $n_{.j}$ est la somme des effectifs de la colonne j et n est la somme totale des effectifs du tableau.



Étape 3 : Calcul de la matrice des distances

La matrice des distances est une matrice carrée qui mesure les distances entre chaque paire de variables. Elle est calculée en utilisant la distance de khi-deux entre chaque paire de variables.

$$d_{\chi^2}^2(i, i') = \sum_{j=1}^J \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i\cdot}} - \frac{f_{i'j}}{f_{i'\cdot}} \right)^2,$$

$$d_{\chi^2}^2(i, G_I) = \sum_{j=1}^J \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i\cdot}} - f_{\cdot j} \right)^2.$$

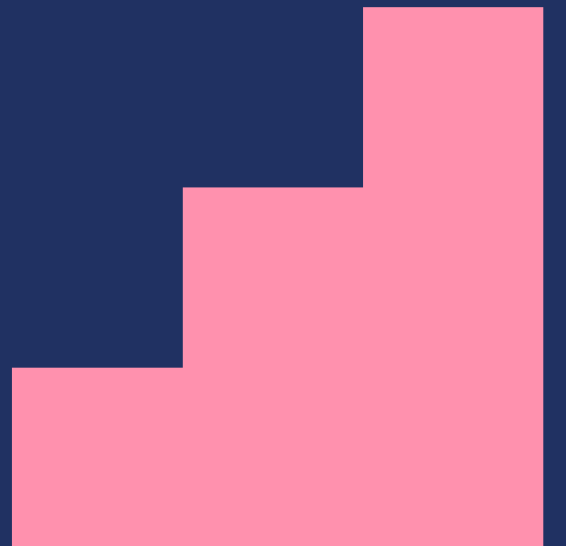


Étape 4 : Calcul des cosinus carrés et des contributions

Les cosinus carrés mesurent l'association entre chaque variable et chaque axe factoriel. Ils sont calculés en utilisant la formule suivante :

$$\text{coscaree}(i, k) = \frac{f_{ik}^2}{\lambda_k \times \sum_{i=1}^I \sum_{k=1}^K f_{ik}^2}$$

où f_{ik} est la fréquence observée de la variable pour la catégorie i , λ_k est la somme des fréquences de la variable et $\sum_{k=1}^K f_{ik}^2$ est la somme des fréquences de la catégorie k .



Les contributions mesurent l'importance de chaque variable pour l'axe factoriel correspondant. Elles sont calculées en utilisant la formule suivante :

$$\textit{contribution}(i, k) = \frac{f_{ik}^2}{\sum_{i=1}^I \sum_{k=1}^K f_{ik}^2}$$

où est la valeur propre de l'axe factoriel . Les contributions sont exprimées en pourcentage et indiquent la part de la variance expliquée par chaque variable sur l'axe factoriel correspondant.



Étape 5 : Calcul des valeurs et vecteurs propres

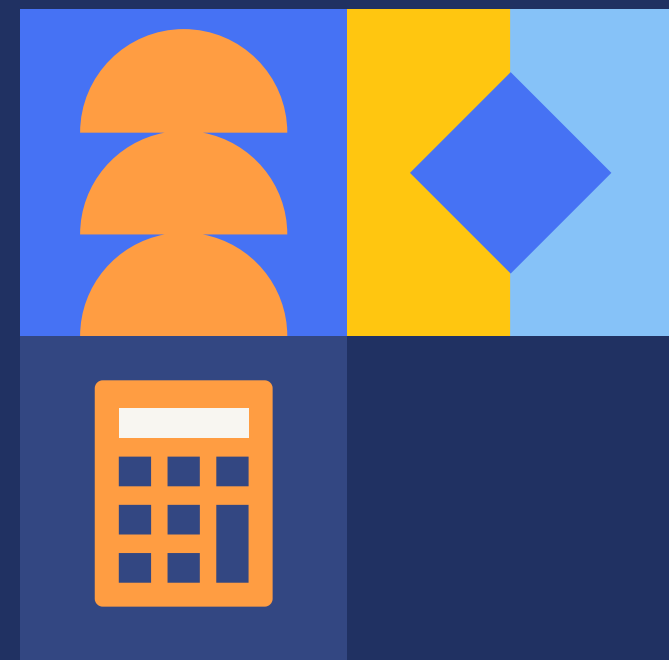
Les valeurs propres sont des mesures de l'importance de chaque axe factoriel pour expliquer la variation totale des données. d'après la matrice des résidus standardisés



Étape 6 : Construction du graphique

Pour construire le graphique de l'AFC, on utilise les coordonnées factorielles des individus et des variables qui ont été calculées à l'étape 4. Les individus sont représentés par des cercles, et les variables sont représentées par des flèches qui indiquent leur contribution à chaque axe factoriel.





avantages et inconvenients de l'AFC





Avantages



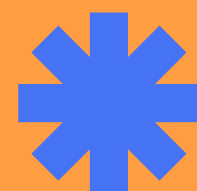
Simplicité

l'AFC facilite
l'analyse des
données
qualitatives.



Réduction

elle réduit
l'information
pour la rendre
plus lisible.



Visualisation

elle permet de
représenter
individus et
modalités ensemble.



Détection

elle aide à trouver
des associations
entre catégories.



Inconvénients



Complexité

difficile à
interpréter si
trop de
données.



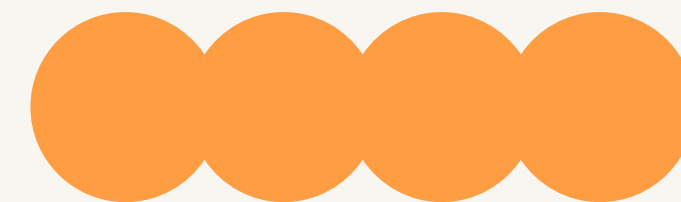
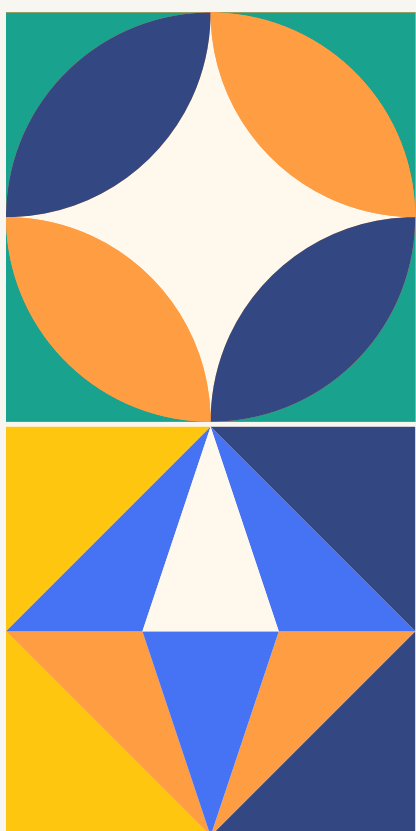
Perte d'info

réduction peut
éliminer des
informations
importantes.

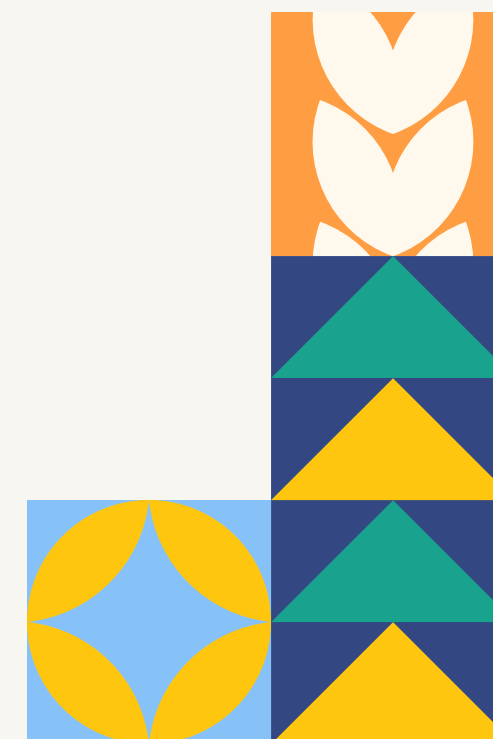


Sensibilité

résultats influencés
par faibles
effectifs.



fin



Presenting Results



Frequency Table

It is used to represent the frequency of different results in the data.

result or
data value



frequency



Color	Number of Friends

Presenting Results

The equivalent frequency of Jan's poll is shown below:

Color	Number of Friends
Pink	6
Green	2
Blue	3
Yellow	4



Which color is the mode?



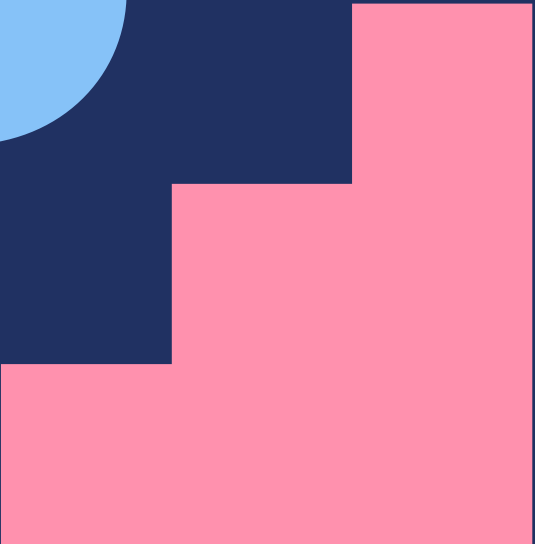
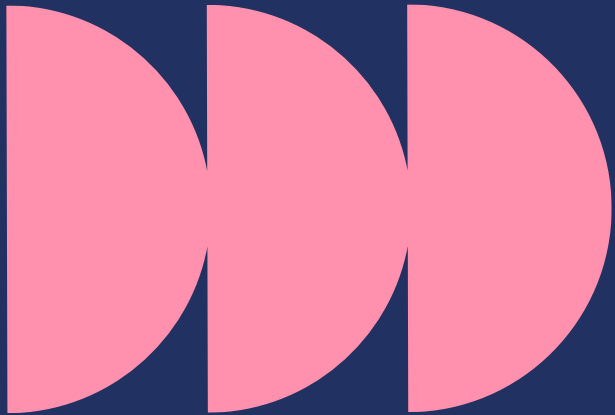
Presenting Results

The equivalent frequency of Jan's poll is shown below:

Color	Number of Friends
Pink	6
Green	2
Blue	3
Yellow	4

Pink appears most often in the poll. Therefore, it is the **mode**.

Remember: A data set may have one mode, multiple modes, or no mode.



Answer Key

Try This!

1. The modes of the data set are 2 and 5.
2. There are 12 students in the class.



modes:
2 and 5

Score	Frequency
2	3
5	3
7	1
12	1
13	2
14	2

Total: 12

Summarizing Results

Two measures of center can be used to summarize numerical data sets: the mean and the median.



Mean

average of all the
data values in the set

$$\text{Mean} = \frac{\text{sum of all data values}}{\text{number of data values}}$$



Median

the "middlemost"
value in the set



Summarizing Results

Find the mean of the following data set: 8, 3, 4, 9, 6

$$\text{Mean} = \frac{\text{sum of all data values}}{\text{number of data values}}$$

$$\text{Mean} = \frac{8 + 3 + 4 + 9 + 6}{5}$$

$$\text{Mean} = \frac{30}{5} = 6$$

The mean or average of the data set is 6.





Summarizing Results

Finding the median changes depending on the number of values in the data set.

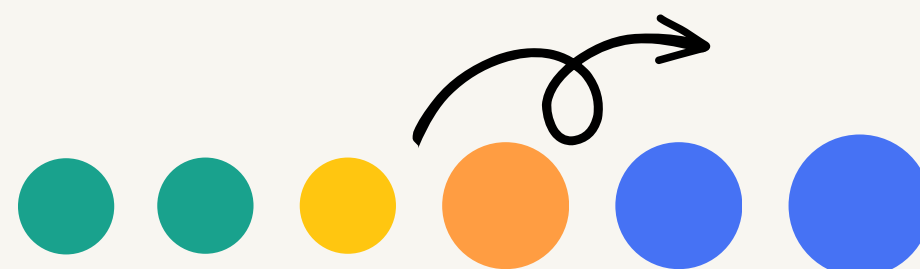
Arrange the values from least to greatest.

least \longrightarrow greatest

If odd number of values:



If even number of values:



Median: $\frac{\text{yellow circle} + \text{orange circle}}{2}$



Summarizing Results

Find the median of the following data sets:

Set A: 8, 3, 4, 9, 6

3, 4, 6, 8, 9



Median

Set B: 11, 17, 3, 14, 19, 7

3, 7, 11, 14, 17, 19

$$\text{Median} = \frac{11 + 14}{2} = 12.5$$

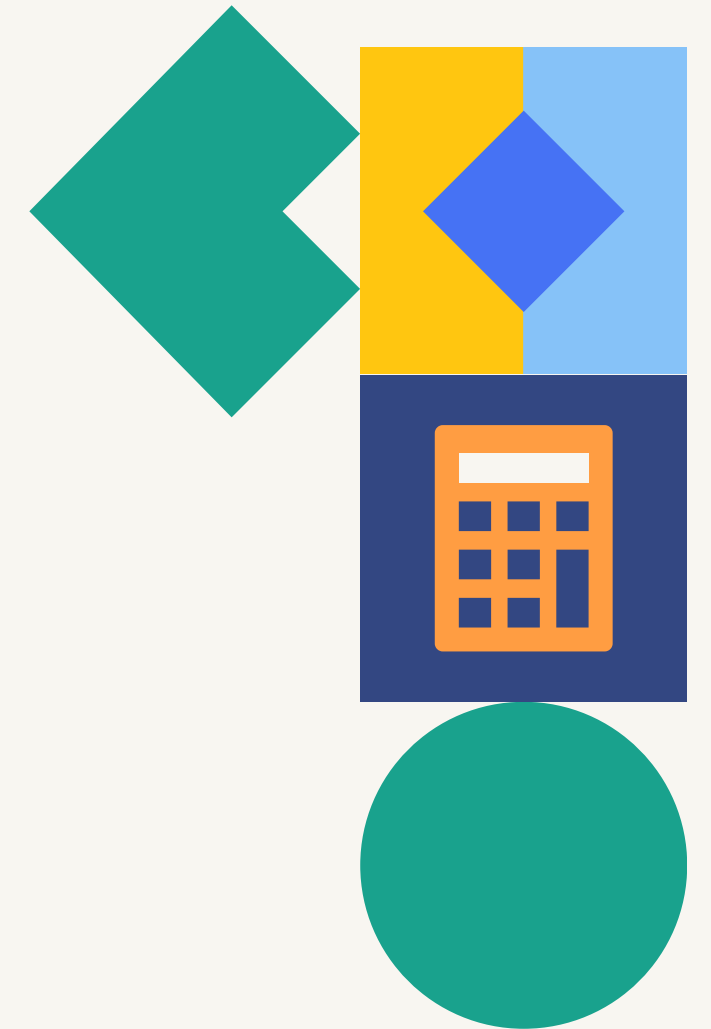


Try This!

Which two data sets have the same mean?

- A 9, 11, 15, 21, 7
- B 12, 11, 14, 15, 9
- C 7, 11, 30, 8, 9
- D 6, 14, 10, 13, 22

Find also the median for all four data sets.



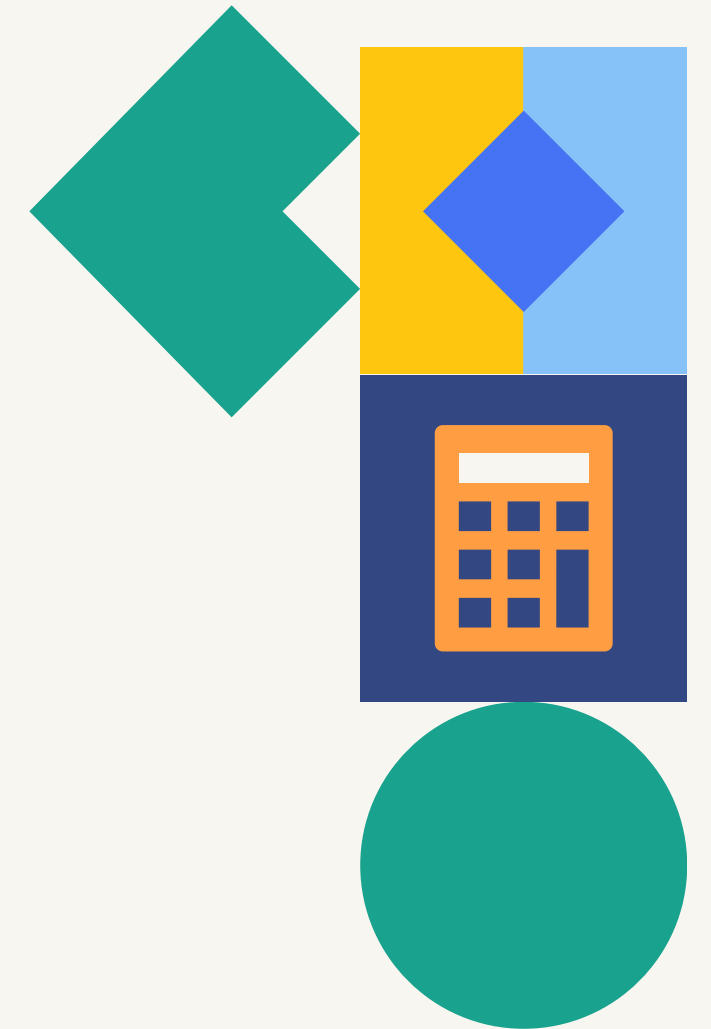
Answer Key

Try This!

Which two data sets have the same mean?

A	9, 11, 15, 21, 7	Mean: 12.6	7, 9, 11, 15, 21	Median: 11
B	12, 11, 14, 15, 9	Mean: 12.2	9, 11, 12, 14, 15	Median: 12
C	7, 11, 30, 8, 9	Mean: 13	7, 8, 9, 11, 30	Median: 9
D	6, 14, 10, 13, 22	Mean: 13	6, 10, 13, 14, 22	Median: 13

C and D have the same mean of 13.



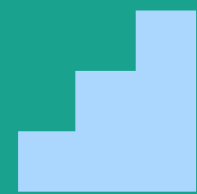


Review



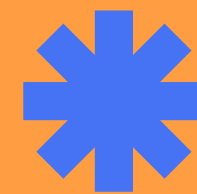
Frequency

number of times
a result appears
in the data set



Mode

result with
the greatest
frequency



Mean

average of all
the data values
in the set



Median

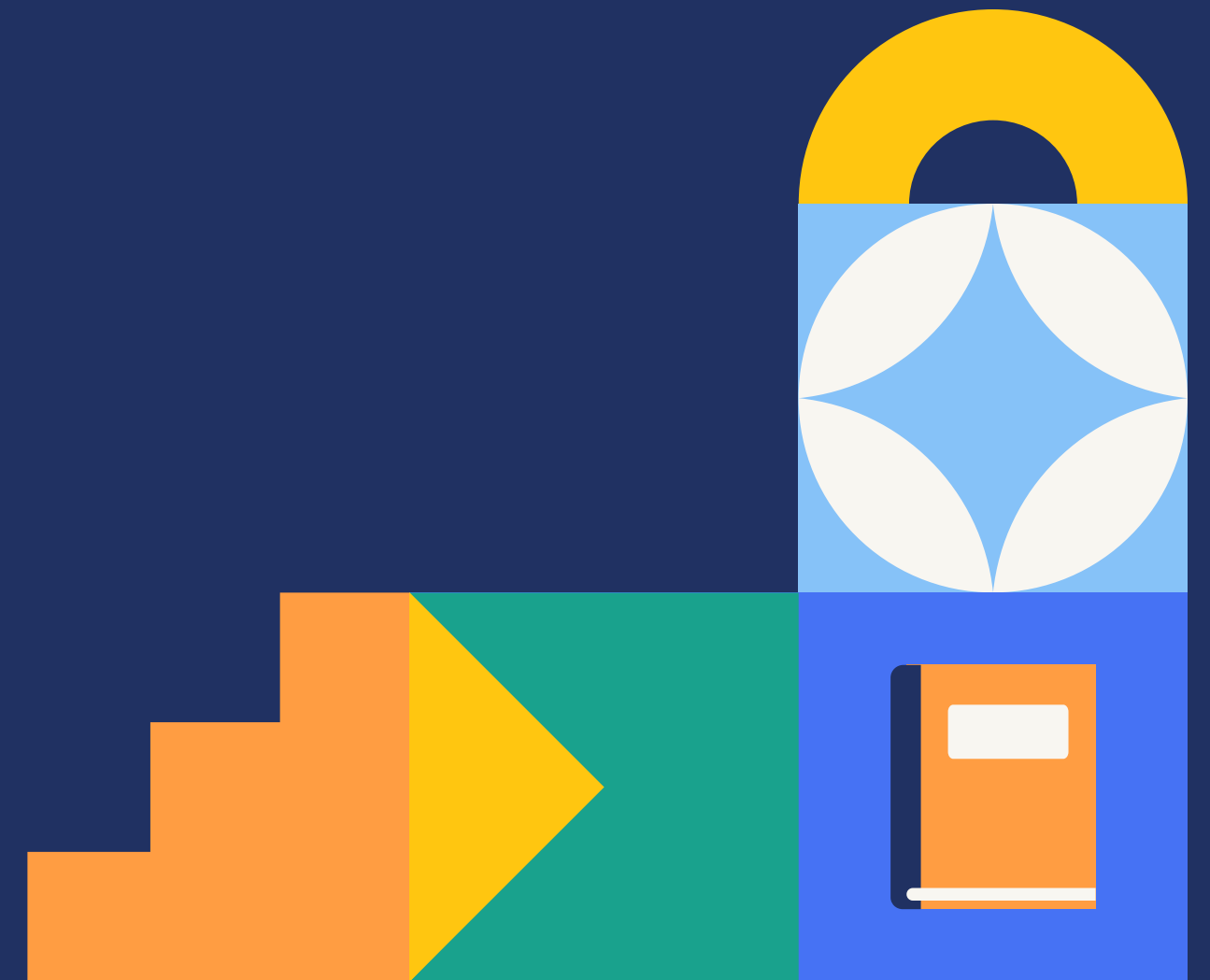
the
"middlemost"
value

Assignment

Design a survey for classmates. Choose one question from the provided options or create your own.

- What are your preferred sports?
- How many pets do you have at home?
- What is your mode of transportation when traveling to school?

Organize, summarize, and present the data you collected in a poster.





Reference

Mathspace. "Statistics." Accessed 14 June 2023,
<https://mathspace.co/textbooks/syllabuses/Syllabus-1155/topics/Topic-21893/subtopics/Subtopic-279788/?searchString=&activeTab=worksheet>.

Resource Page

Use these icons and illustrations in your Canva Presentation. Happy designing! Don't forget to delete this page before presenting.





Try this background for online class.
*Please delete this section before downloading.



**Press these
keys while
on Present
mode!**



for blur



for a drumroll



for bubbles



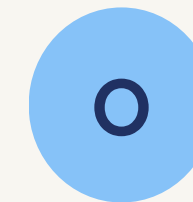
for unveil



for confetti



for mic drop



for quiet



Any number from
0-9 for a timer