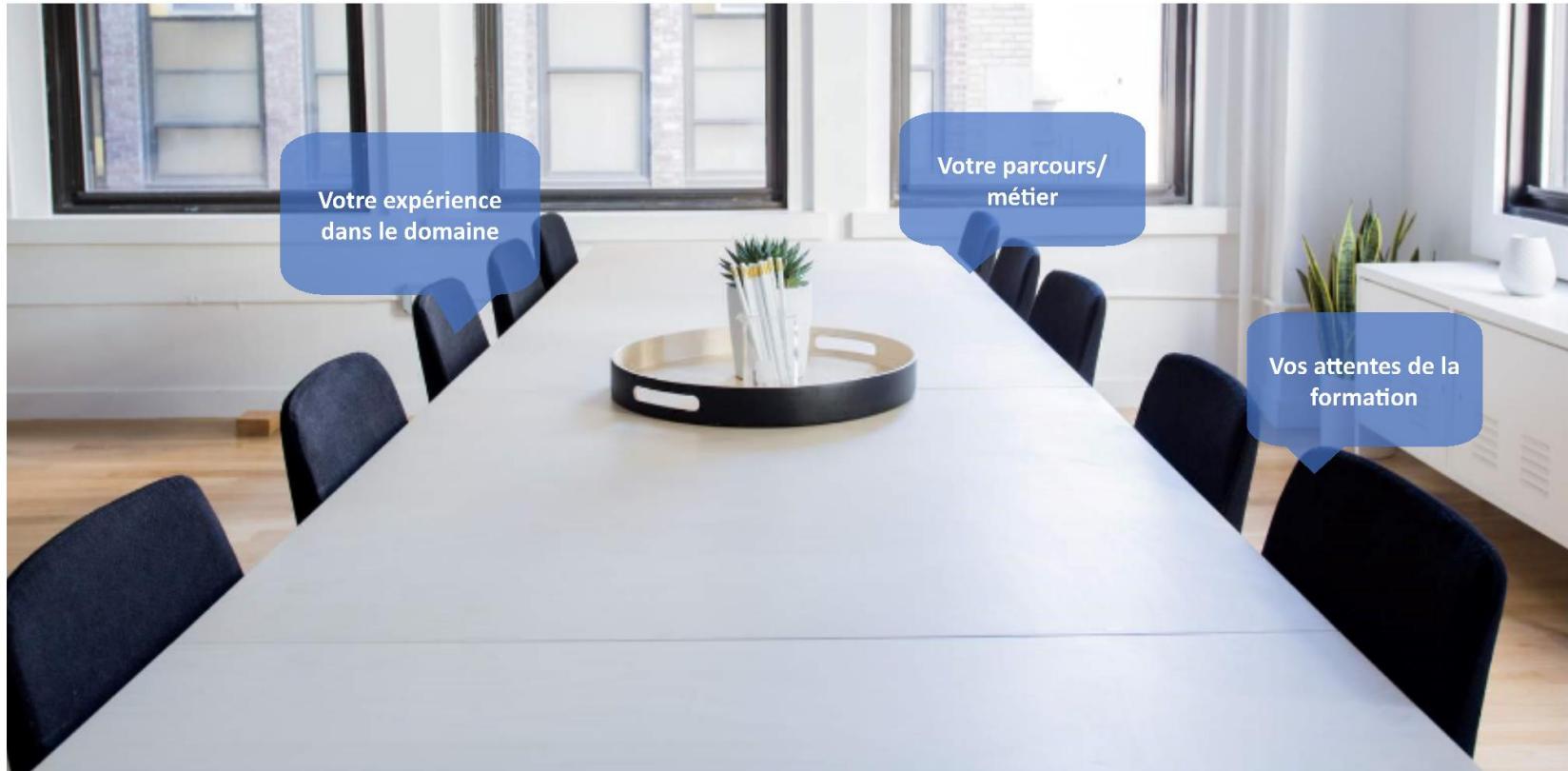




Rendre l'invisible visible avec la Data Visualisation

Votre formateur : Gaoussou DIATTA

Faisons connaissance !



Objectifs de la formation

- Se familiariser avec les différentes techniques de représentation des données et d'amélioration de la lisibilité
- Connaître les avantages et inconvénients de chaque technique de représentation
- Savoir sélectionner le(s) visuel(s) adapté(s) pour améliorer la compréhension
- Pratiquer sur plusieurs exemples



Rendre l'invisible visible avec la Data
Visualisation

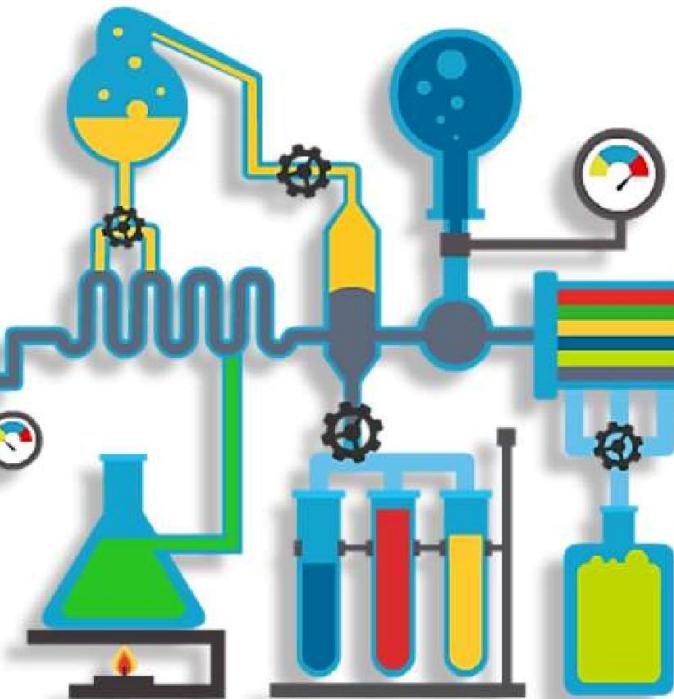
Data visualisation : Fondamentaux de la data visualisation

- Qu'est-ce que la Data visualisation ?

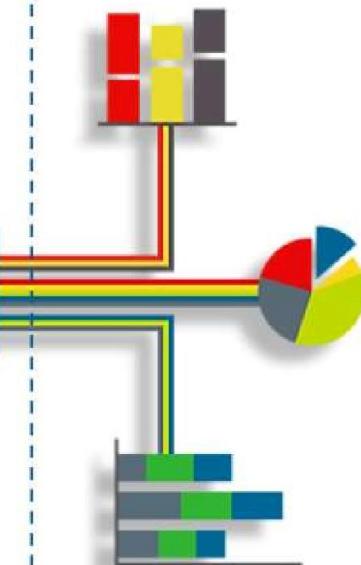
Integrate Different Data Sets



Analyze



Visualize



Data visualisation : Fondamentaux de la data visualisation

De quelles manières peut-on présenter de l'information ?

- **Texte** : l'information est présentée sous forme de phrases, paragraphes, et sections. Cela nécessite un travail de réflexion pour présenter l'information sous une claire. S'il s'agit de présenter un petit nombre de statistiques (jusqu'à cinq), cette forme peut être adaptée,
- **Tableau** : il s'agit d'une présentation systématique d'informations statistiques (quantitatives et/ou qualitatives) sous la forme de lignes et de colonnes incluant leur description et aussi un titre du tableau. L'information y est présentée sans emphase particulière sur tel ou tel aspect des données : le lecteur choisit lui-même l'information particulière qui l'intéresse. Contrairement à un graphique, on peut y voir l'information avec une certaine précision (e.g. nombreuses décimales),
- **Graphique** : permet de simplifier et de synthétiser une information complexe à travers des éléments visuels (histogramme, camembert...), comme par exemple des proportions ou des tendances. Le choix du type de graphique doit se faire en fonction de ce qu'on souhaite porter à l'attention du lecteur (*nudging*).

Nota : le cerveau humain peut mieux comprendre des motifs (patterns), des différences, ou des relations à travers des représentations visuelles.

Data visualisation : Fondamentaux de la data visualisation

- Pour créer des visuels efficaces :
 - Comprendre le contexte et le(s) public(s) cible(s),
 - Déterminer le(s) graphique(s) adapté(s),
 - Identifier et éliminer les éléments qui encombrent votre graphique (bruit, informations inutiles),
 - Focaliser l'attention de votre public (cf. *nudging*) sur les éléments des données qui sont vraiment importants,
 - Se mettre dans la peau d'un graphiste pour utiliser les outils disponibles,
 - Faire en sorte que le message que vous souhaitez passer parle à votre public : faire du *story telling*.

Data visualisation : Fondamentaux de la data visualisation

- Le quartet d'Anscombe :

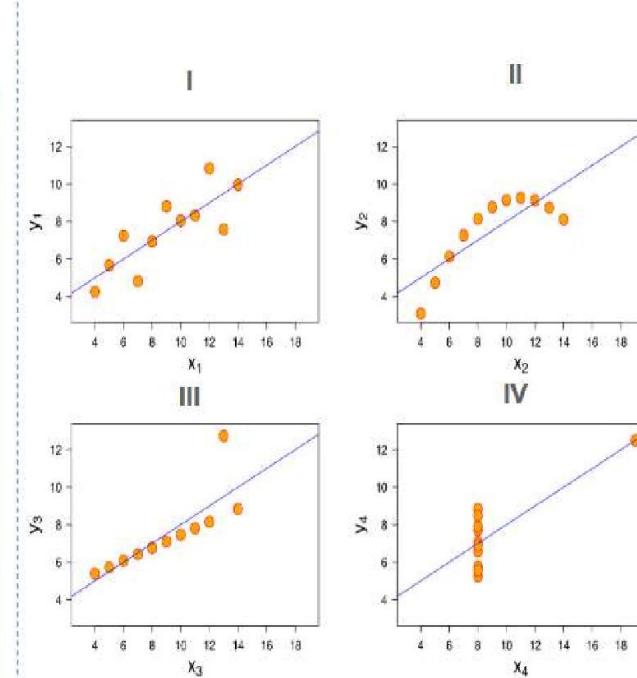
I	
X	Y
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68
SUM	99.0
AVG	9.00
STDEV	3.32

II	
X	Y
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.13
7	7.26
5	4.74
SUM	99.0
AVG	9.00
STDEV	3.32

III	
X	Y
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73
SUM	99.0
AVG	9.00
STDEV	3.32

IV	
X	Y
8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89
SUM	99.0
AVG	9.00
STDEV	3.32

Anscombe's Quartet Data - Set

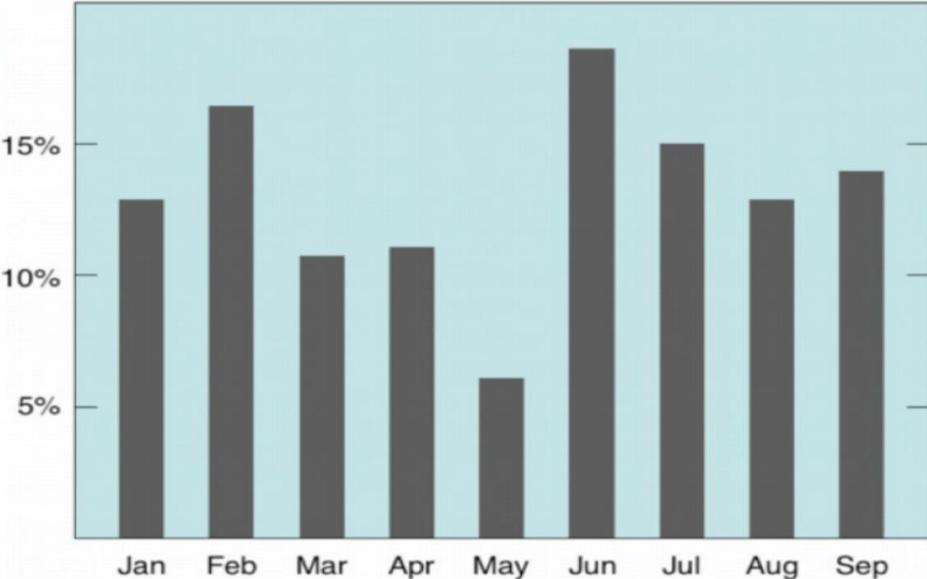
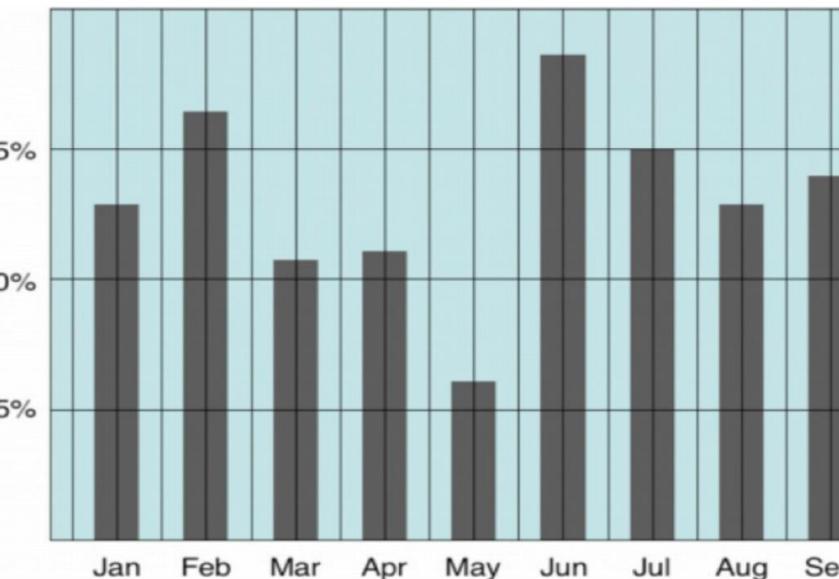


Graphical Data Visualization

Data visualisation : Fondamentaux de la data visualisation

- Clarté et simplicité : exemple avec le *data-to-ink ratio*

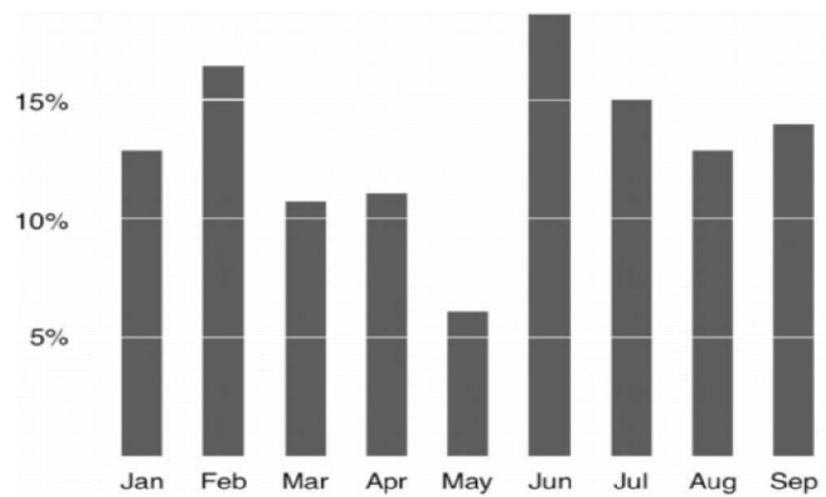
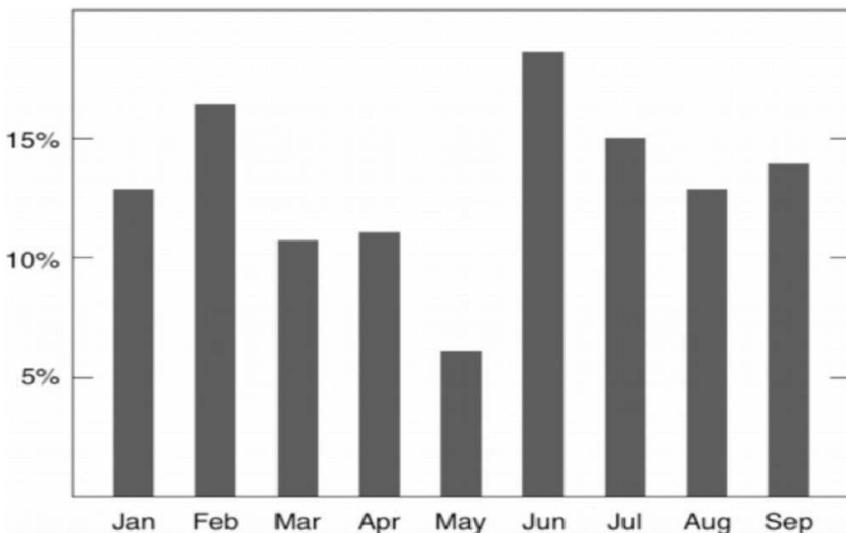
Data-ink ratio = $\frac{\text{Data-ink}}{\text{Total ink used}}$.



Data visualisation : Fondamentaux de la data visualisation

- Clarté et simplicité : exemple avec le *data-to-ink ratio*

Data-ink ratio = $\frac{\text{Data-ink}}{\text{Total ink used}}$.



Data visualisation : Fondamentaux de la data visualisation

- La pré-attention : le cerveau commence à traiter les données qu'il reçoit même sans intention de notre part. Démonstration :

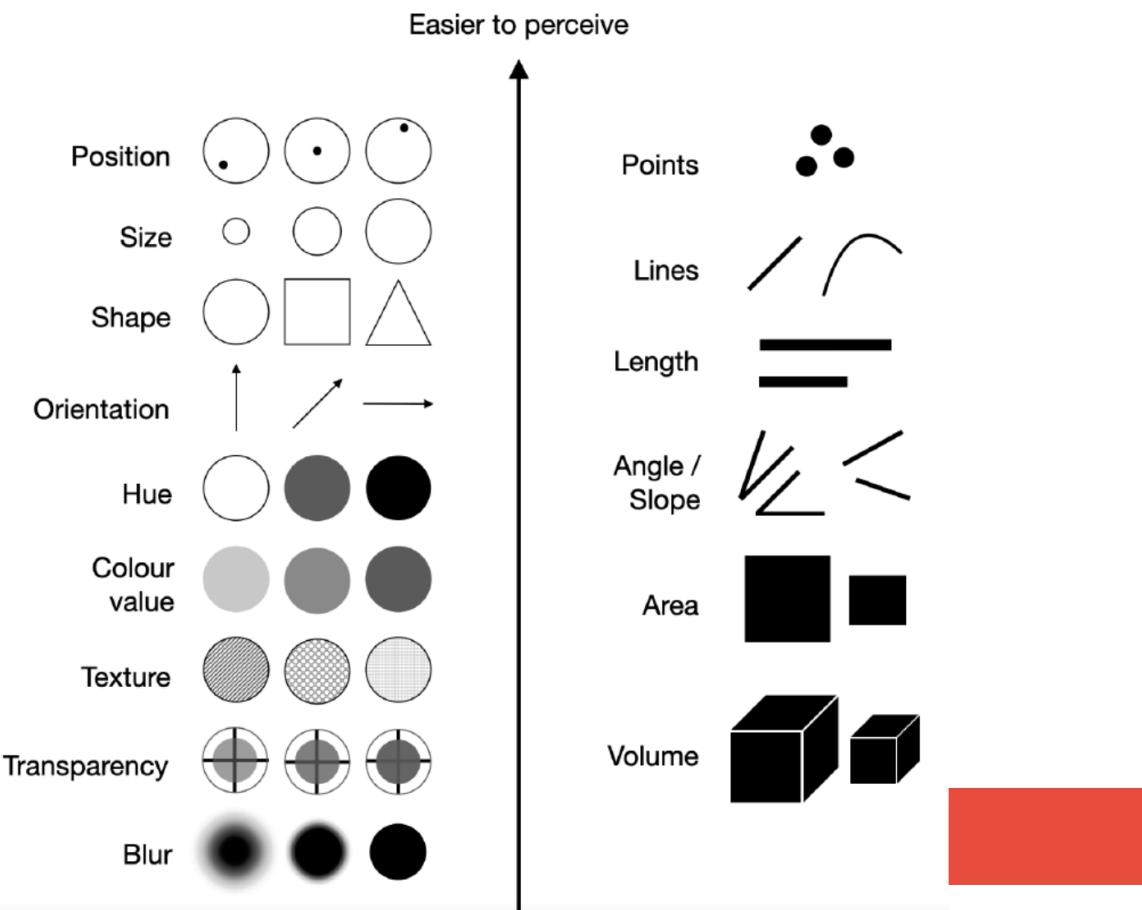
1) Un(e) participant(e) lira à haute voix les mots suivants:

ROUGE VERT BLEU JAUNE ROUGE VERT BLEU JAUNE
ROUGE VERT BLEU JAUNE ROUGE VERT BLEU JAUNE

2) Ensuite, dire la couleur de chaque mot.

Data visualisation : Fondamentaux de la data visualisation

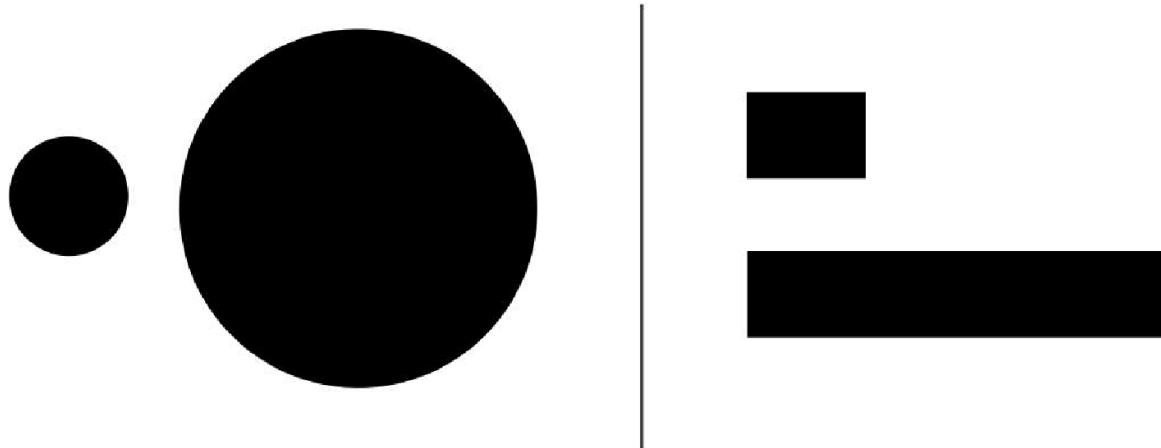
- La pré-attention : quelles présentations d'information sont-elles plus faciles à saisir par le cerveau ?



Data visualisation : Fondamentaux de la data visualisation

- La pré-attention : quelles présentations d'information sont-elles plus faciles à saisir par le cerveau ?

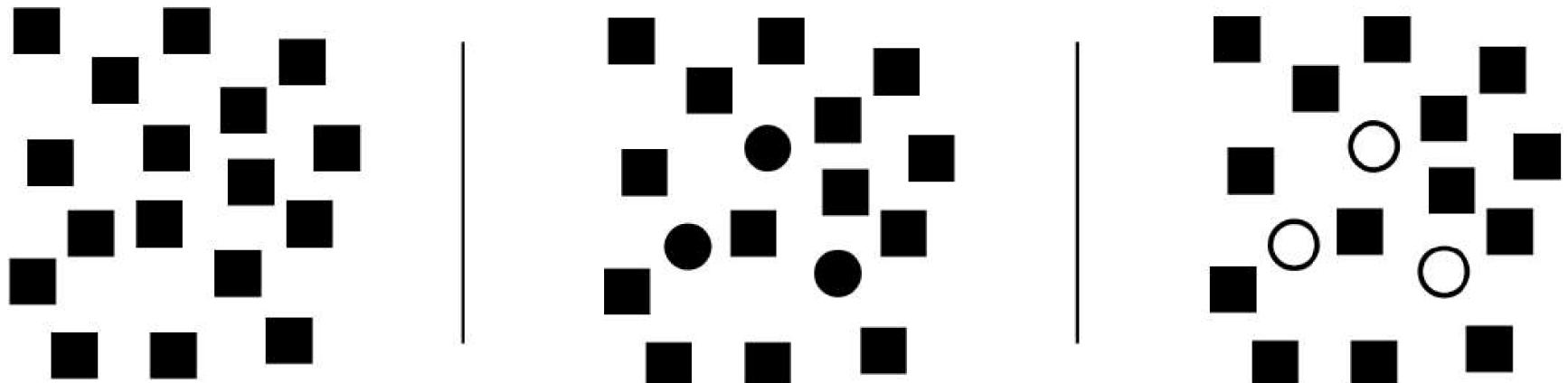
Exemple : on veut relayer précisément la différence de taille



Data visualisation : Fondamentaux de la data visualisation

- La pré-attention : quelles présentations d'information sont-elles plus faciles à saisir par le cerveau ?

Exemple : on veut relayer précisément la différence



Data visualisation : Fondamentaux de la data visualisation

- Comment choisir les variables de représentation en fonction du type de données ?

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density (Color Value)	Color Hue
Angle	Color Saturation	Texture
Slope	Color Hue	Connection
Area	Texture	Containment
Volume	Connection	Density (Color Value)
Density (Color Value)	Containment	Color Saturation
Color Saturation	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area	Slope
Containment	Volume	Area
Shape	Shape	Volume

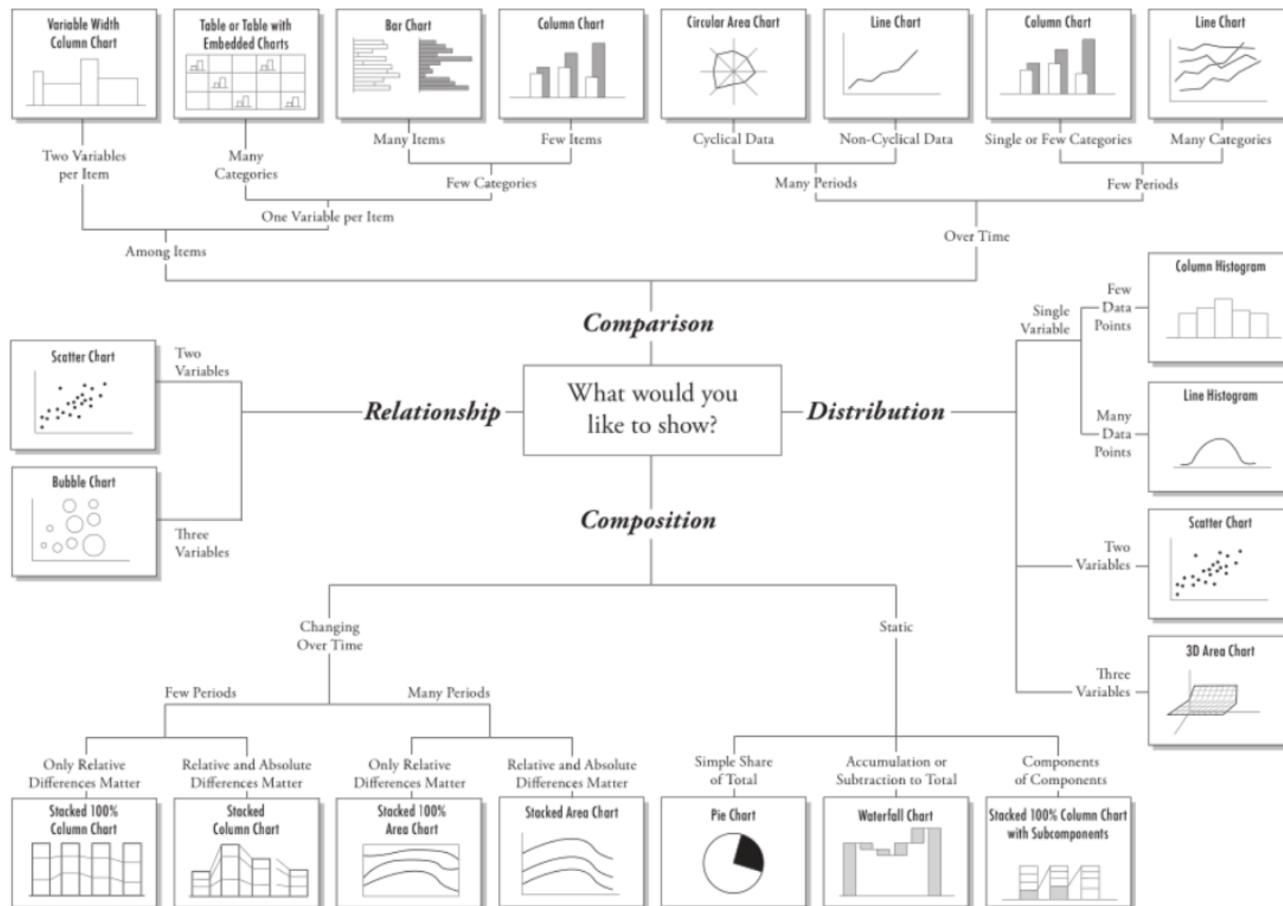
Data visualisation : Fondamentaux de la data visualisation

- **Comment choisir le bon graphique en fonction du type de données ?**

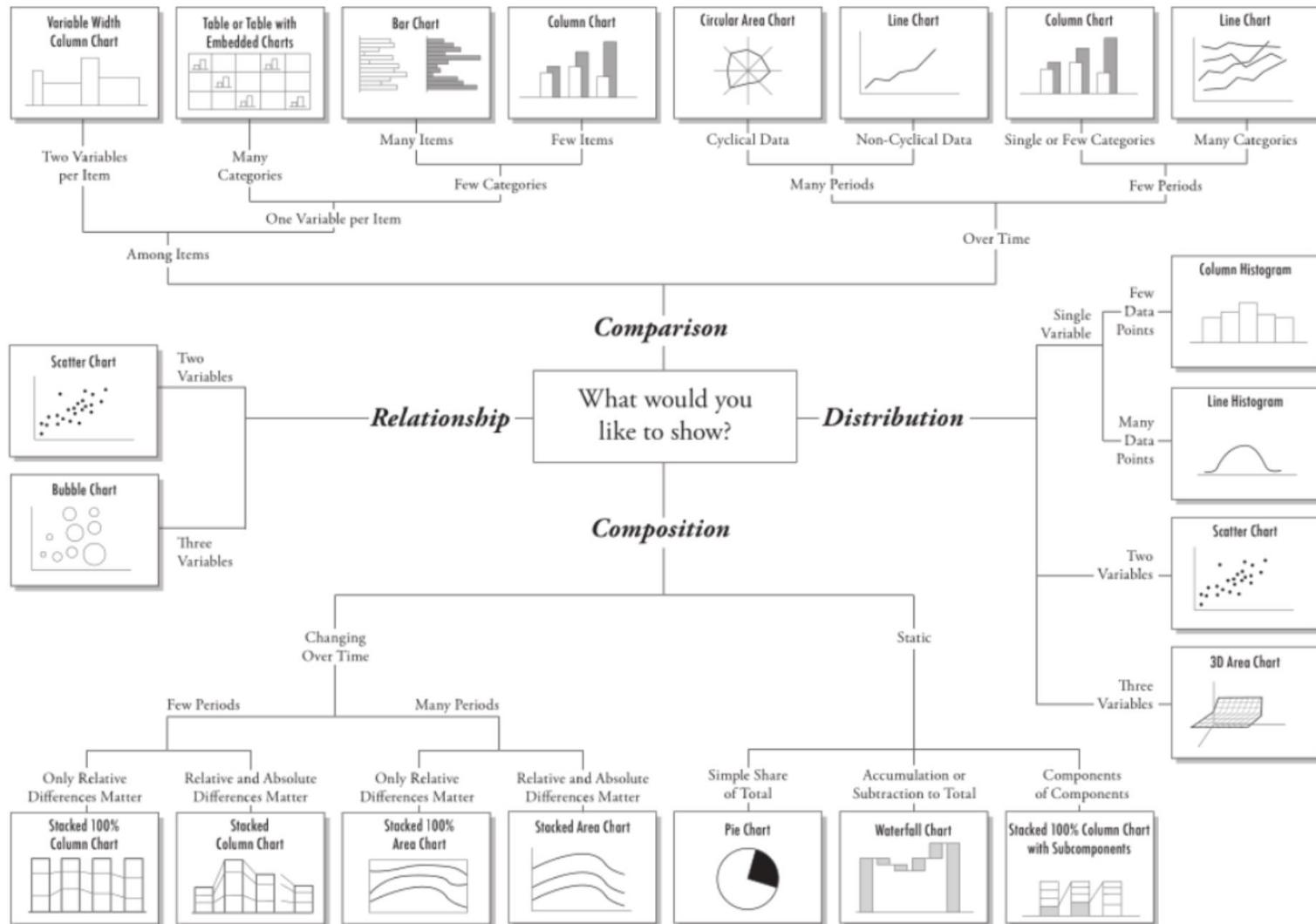
Question	Aim	Data Types	Chart
How is the data distributed?	Distribution	1 Continuous	Histogram, Box plot, Violin plot
How often does each value appear?	Comparison	1 Categorical	Bar chart, Pie chart
How do the parts relate to the whole?	Composition	1 Categorical	Bar chart, Stacked bar chart
Is there a trend or pattern?	Trend	1 Continuous & 1 Ordinal	Line chart, Area chart
Is there a correlation between variables?	Relationship	2 Continuous	Scatterplot, Bubble chart
How often do certain values occur?	Comparison	2 Categorical	Heatmap
Are the distributions similar in each grouping?	Distribution & Comparison	1 Continuous & 1 Categorical	Small multiples of 1 continuous with a categorical encoding (e.g. colour)

Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ?



Data visualisation : Fondamentaux de la data visualisation



Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ?
- Quels sont les avantages/inconvénients de chaque graphique ?
- Lequel choisiriez-vous ?

OPTION A: pies

"I plan to be working here in one year"

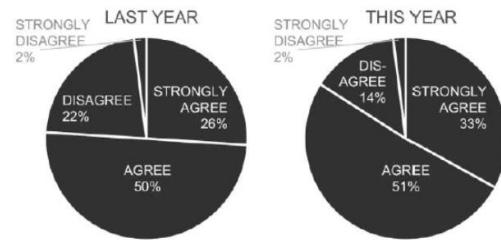
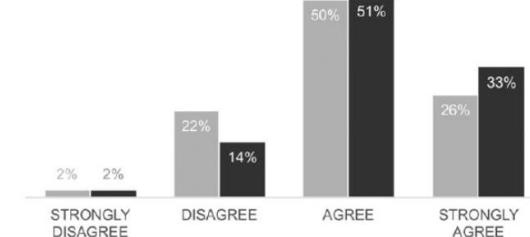


FIGURE 2.12a Pies

OPTION B: bars

"I plan to be working here in one year"

LAST YEAR | THIS YEAR



19

OPTION C: divergent stacked bars

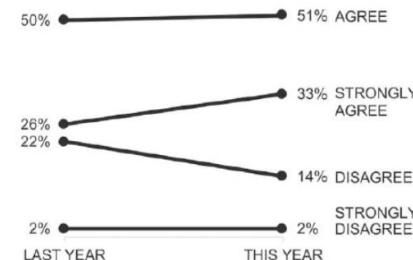
"I plan to be working here in one year"



FIGURE 2.12c Divergent stacked bars

OPTION D: slopegraph

"I plan to be working here in one year"



Data visualisation : Fondamentaux de la data visualisation

- **Comment choisir un graphique ? Exercice 1**

Explorez le jeu de données qui se trouve [ici](#).

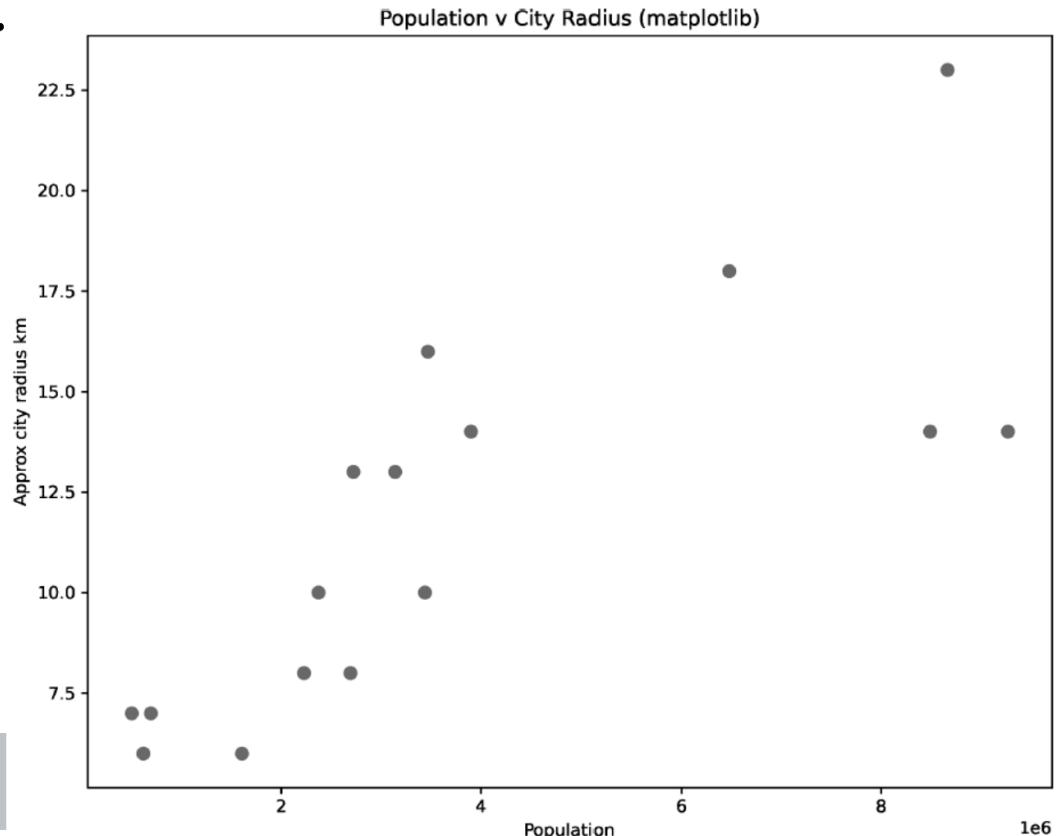
- Nous souhaitons visualiser le lien entre la taille d'une ville, et la population qu'elle contient. Que proposeriez-vous comme graphique ?

- Maintenant nous souhaitons aussi faire apparaître dans le même graphique l'information sur la taille de la ville ('attribut 'City Size'), ainsi que l'information de 'People per dwelling'.

Proposez une façon de faire.

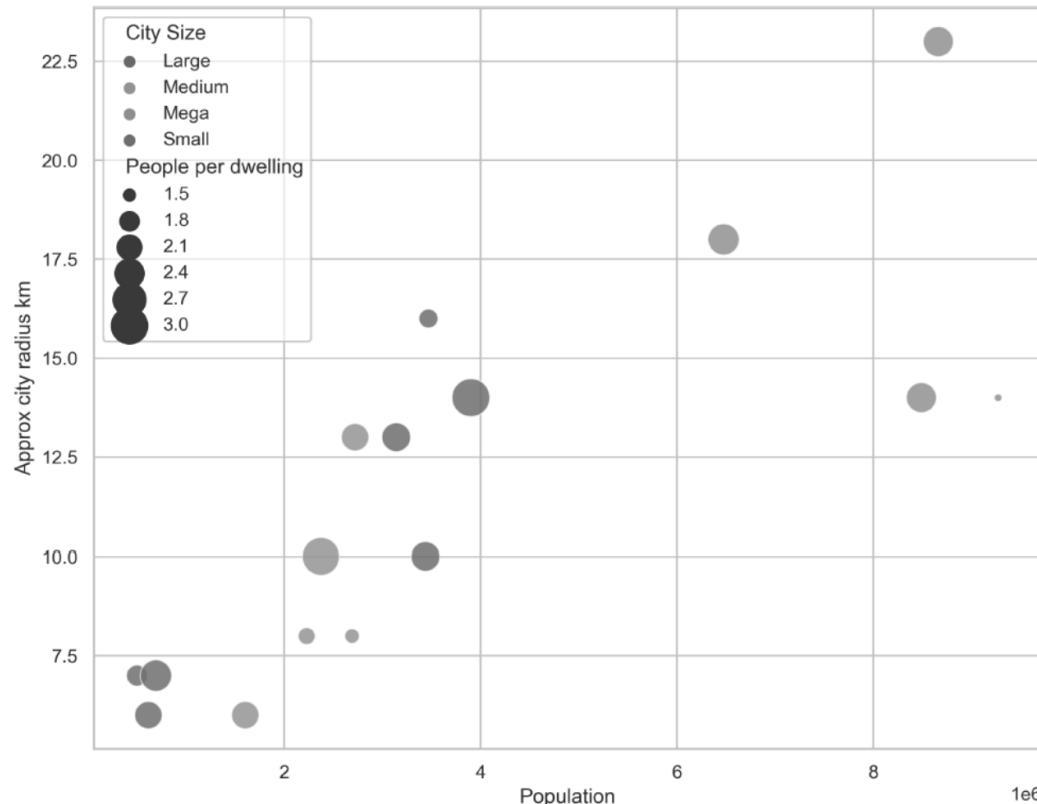
Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 1
 - Nous souhaitons visualiser le lien entre la taille d'une ville, et la population qu'elle contient.



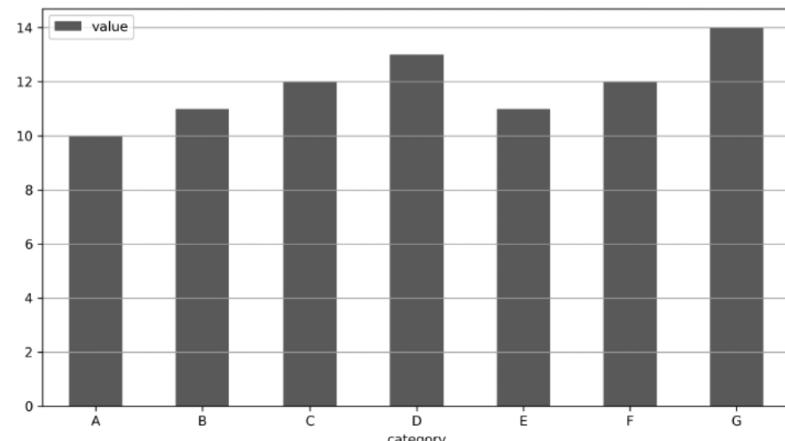
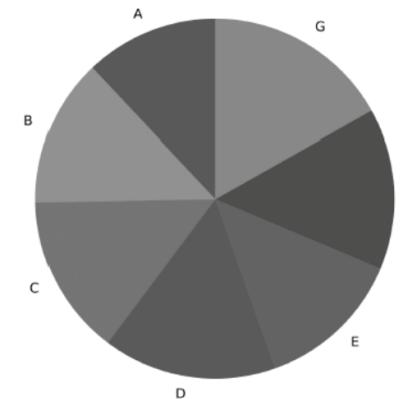
Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 1
 - Maintenant nous souhaitons aussi faire apparaître dans le même graphique l'information sur la taille de la ville (attribut 'City Size'), ainsi que l'information de 'People per dwelling'.



Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 2
 - Quels avantages/inconvénients de chaque alternative pour comparer les tailles des différents groupes ?
 - Quel est le meilleur graphique ?

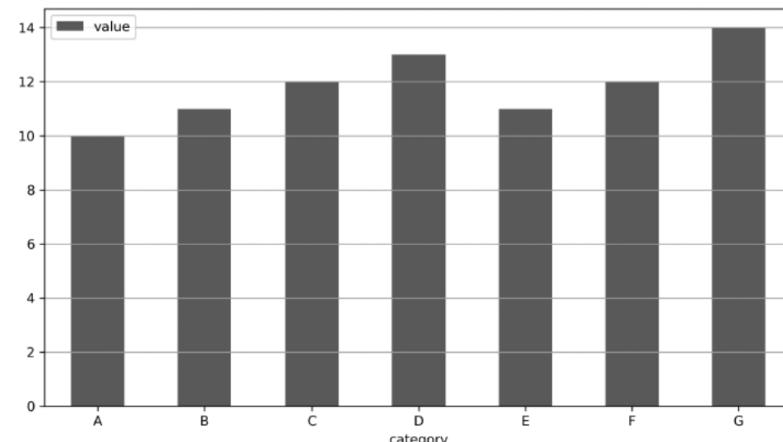
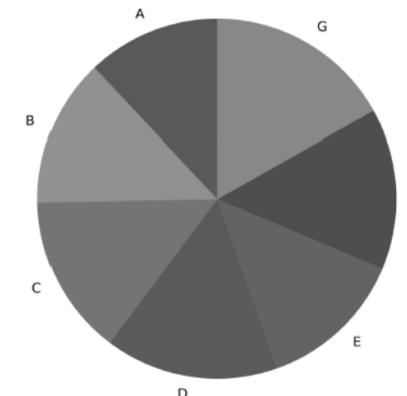
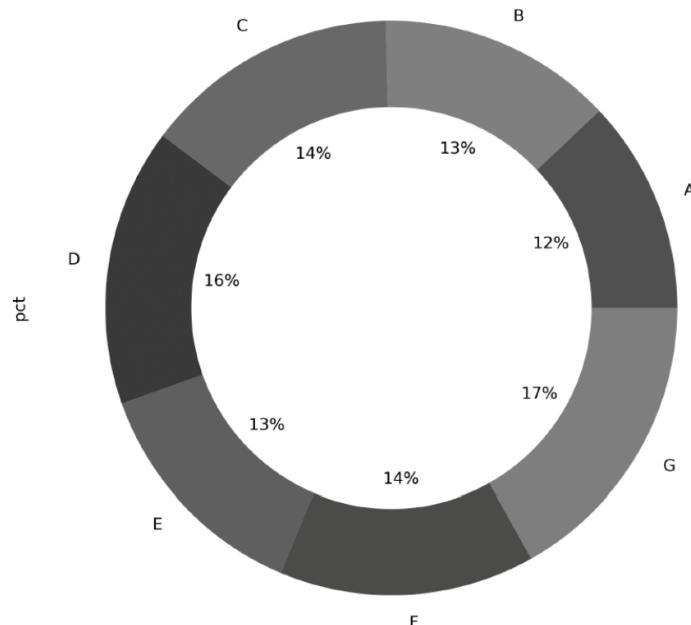


Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 2

- Une alternative qui est compacte

ET lisible : le donut chart.



Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 3
 - Comment peut-on améliorer le tableau ci-dessous ?
 - Quel type de graphique pourrait le remplacer avantageusement ?

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

New client tier share

TIER	ACCOUNTS		REVENUE	
	#	% OF TOT	\$M	% OF TOT
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

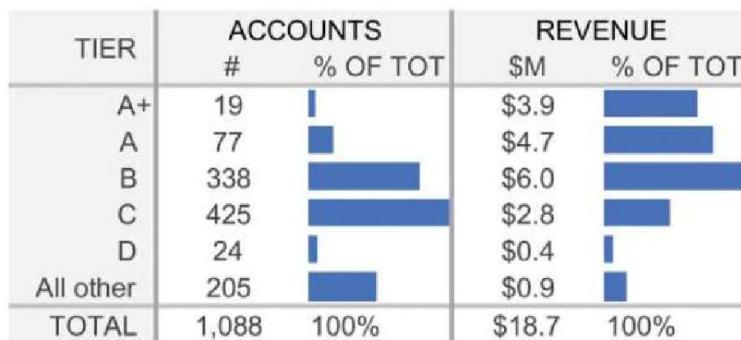
Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

New client tier share



Data visualisation : Fondamentaux de la data visualisation

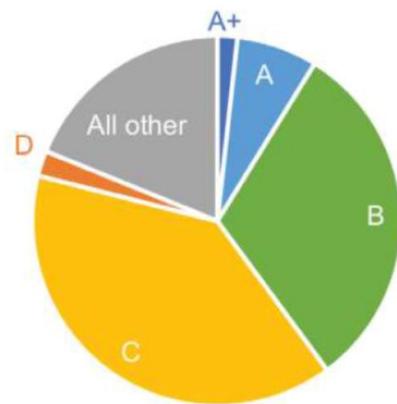
• Comment choisir un graphique ? Exercice 3

New client tier share

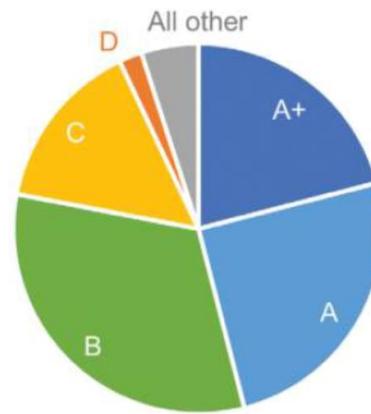
Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

New client tier share

% of Total Accounts



% of Total Revenue

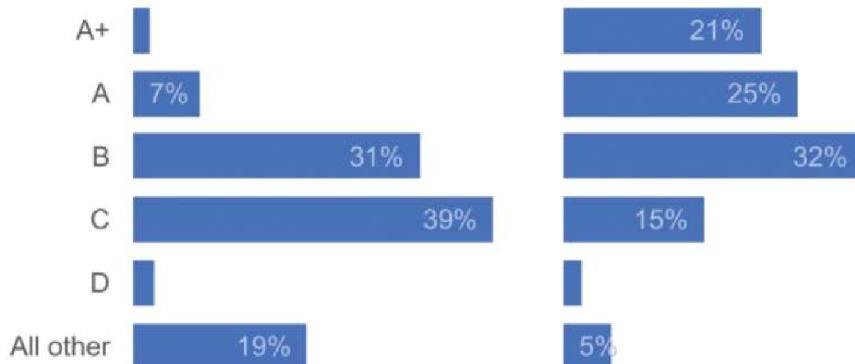


Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share

TIER | % OF TOTAL ACCOUNTS | % OF TOTAL REVENUE



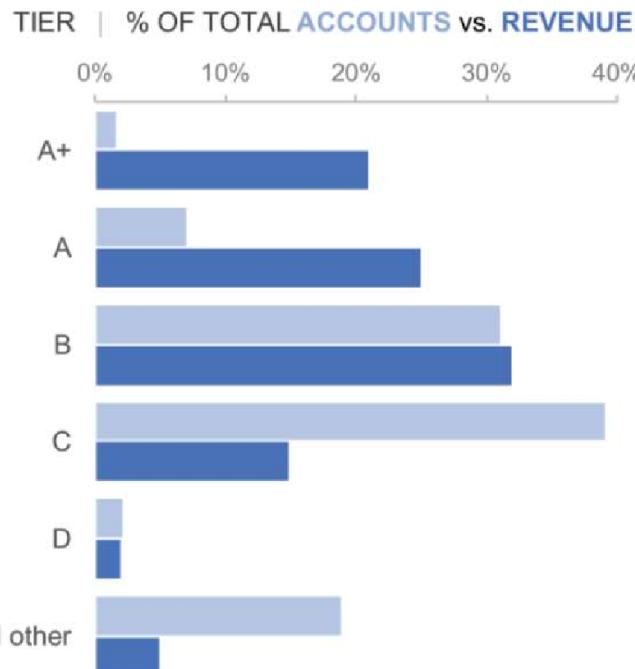
New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share



New client tier share

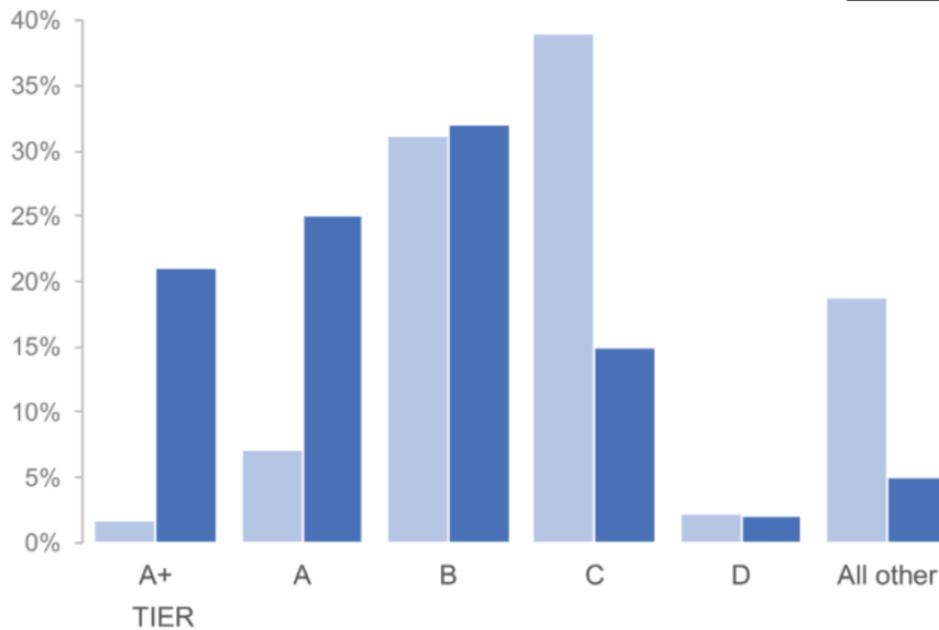
Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE



New client tier share

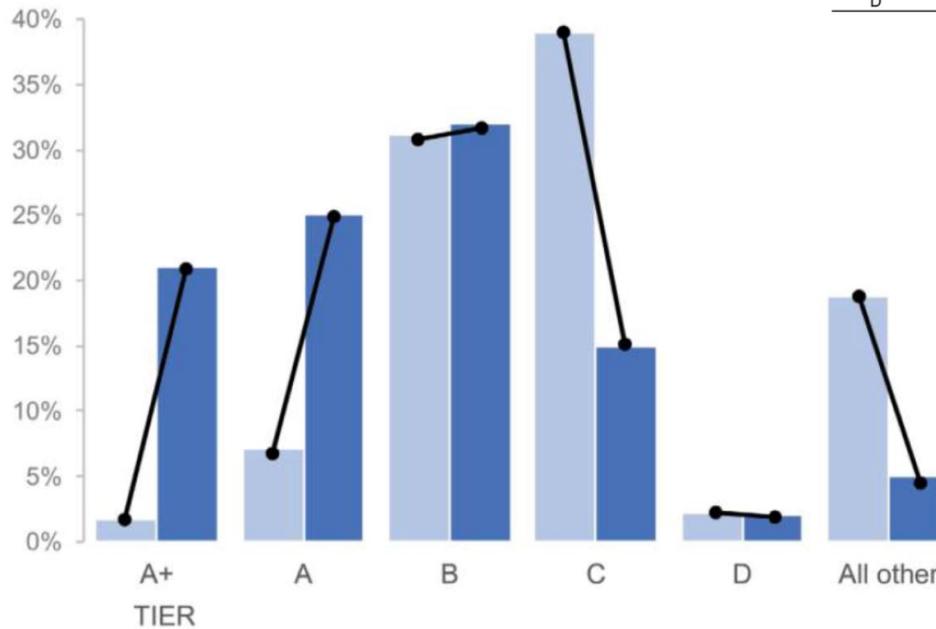
Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 3

New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE



New client tier share

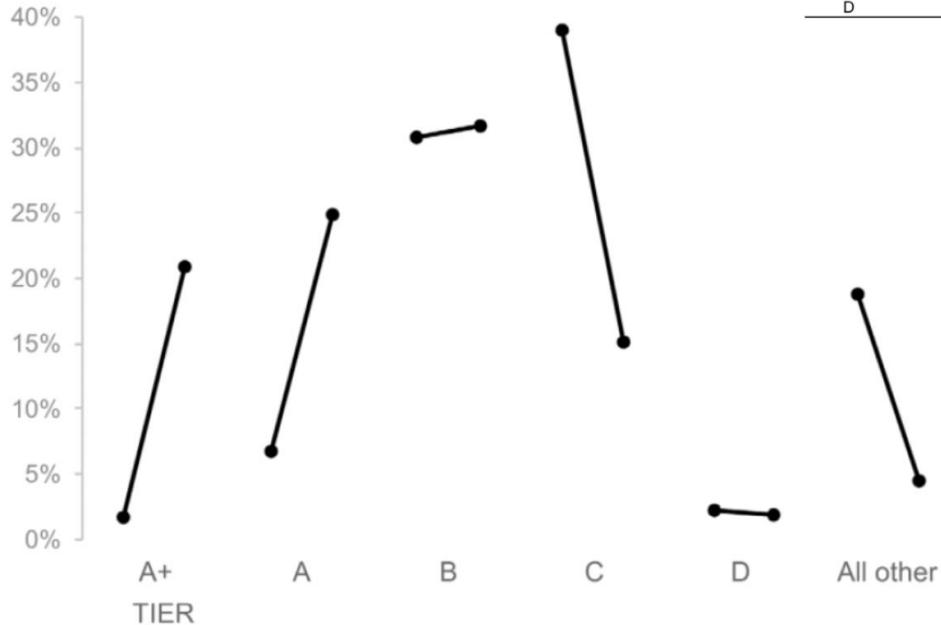
Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE



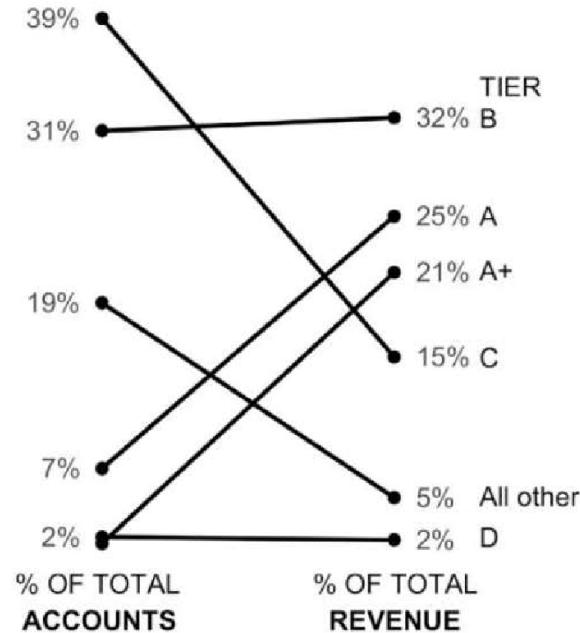
New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 3

New client tier share



New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 4
- Comment peut-on améliorer le tableau ci-dessous ?

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 4

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

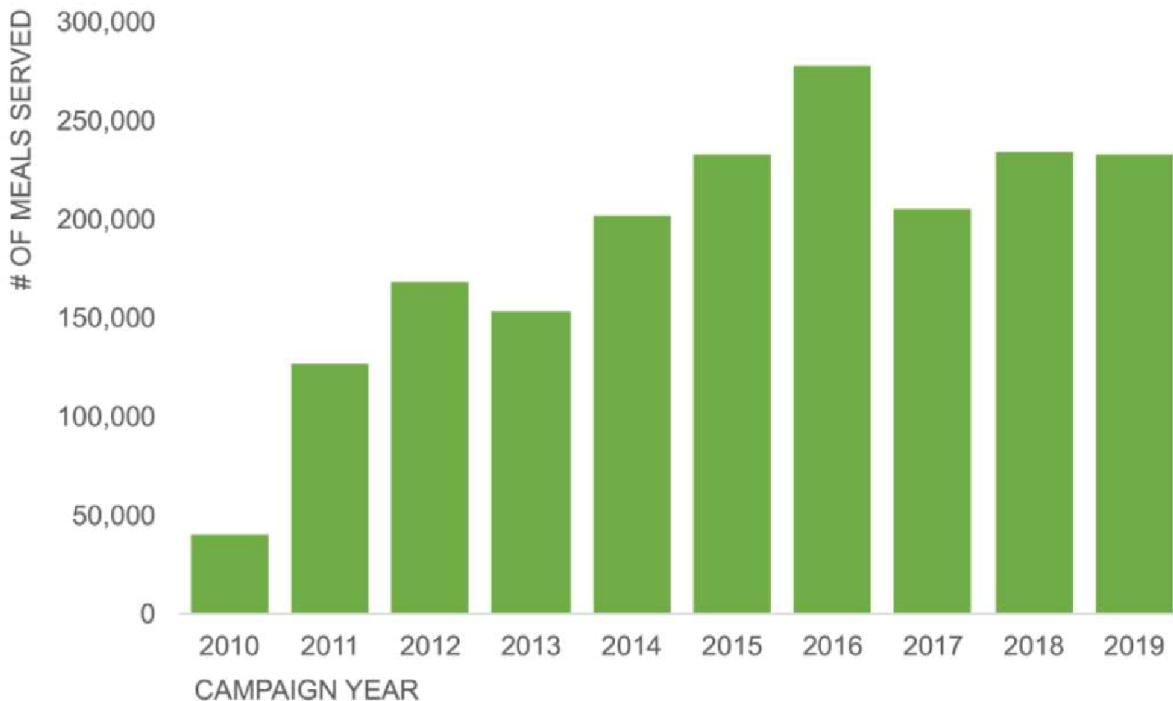
Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

Data visualisation : Fondamentaux de la data visualisation

- **Comment choisir un graphique ? Exercice 4**

Meals served over time



Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 4

Meals served over time

OF MEALS SERVED

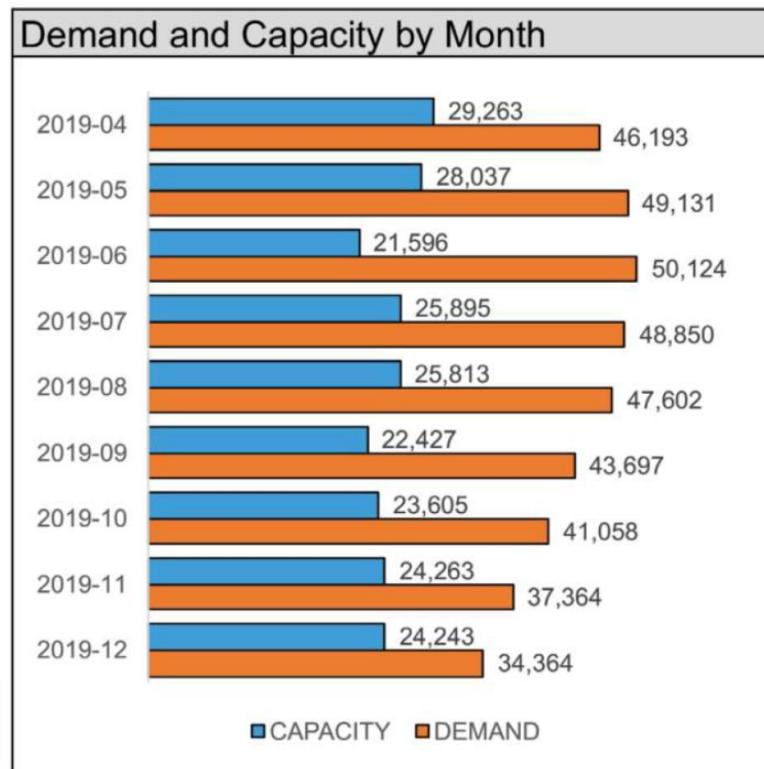


Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 5
- Comment peut-on améliorer le graphique ci-dessous ?

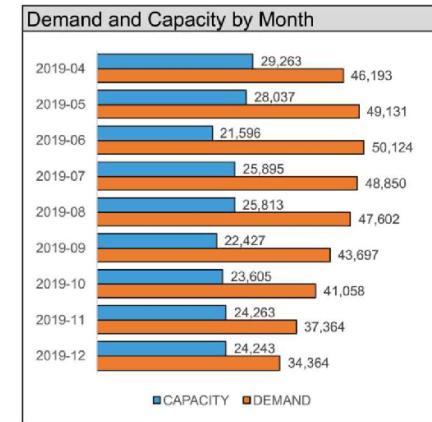
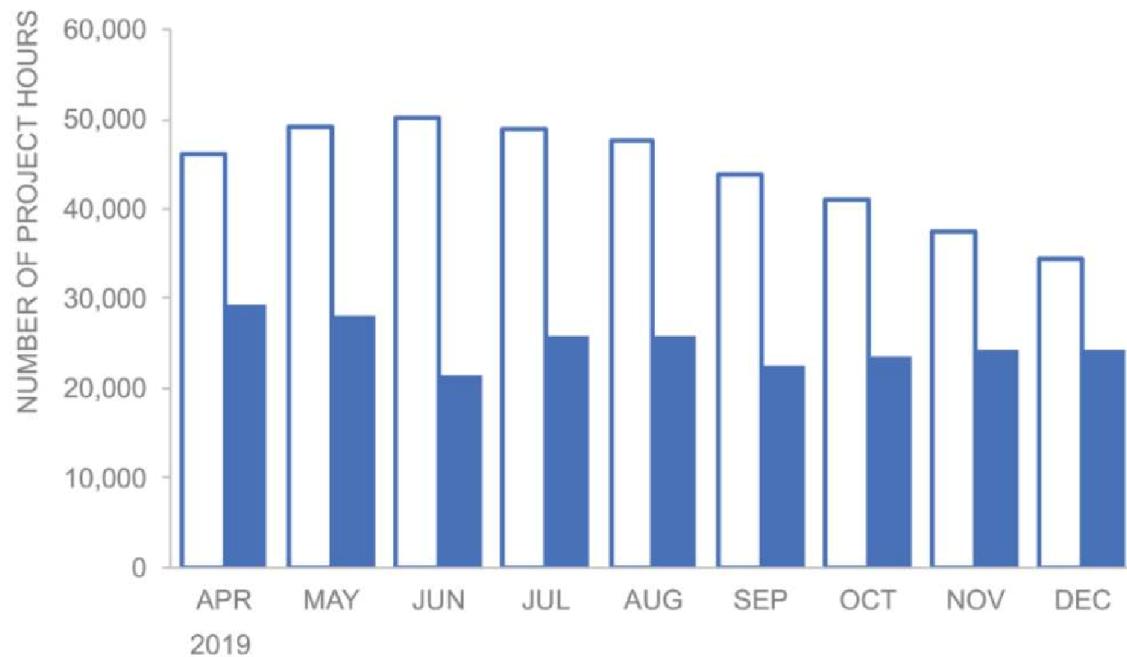


Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 5

Demand vs capacity over time

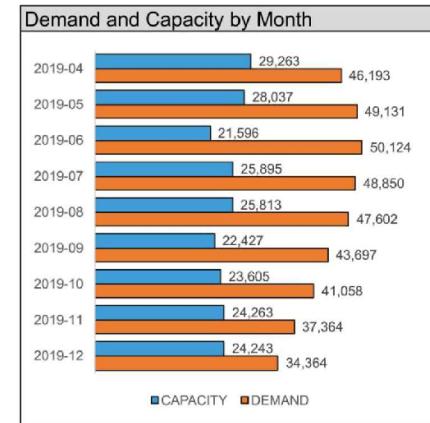
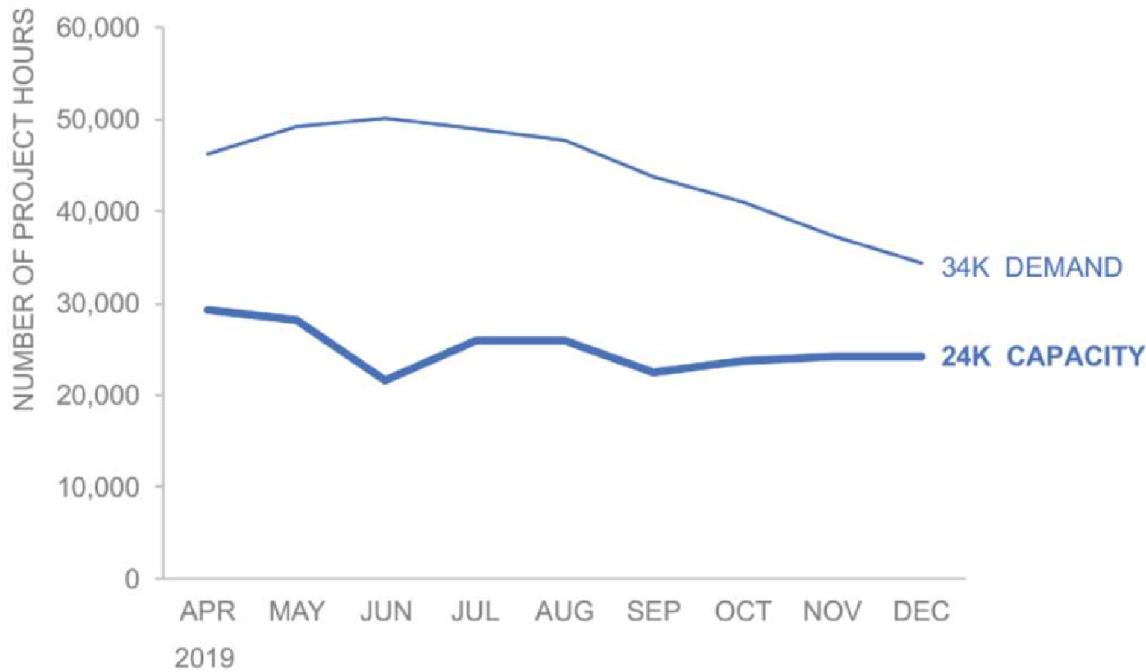
DEMAND | CAPACITY



Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 5

Demand vs capacity over time

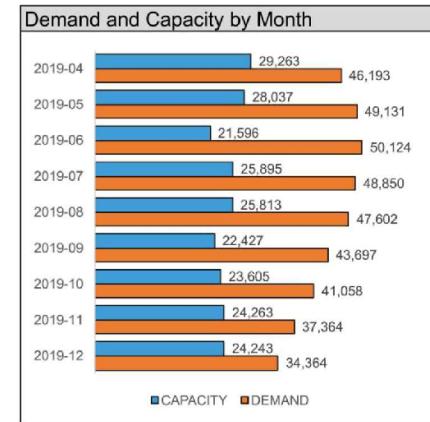
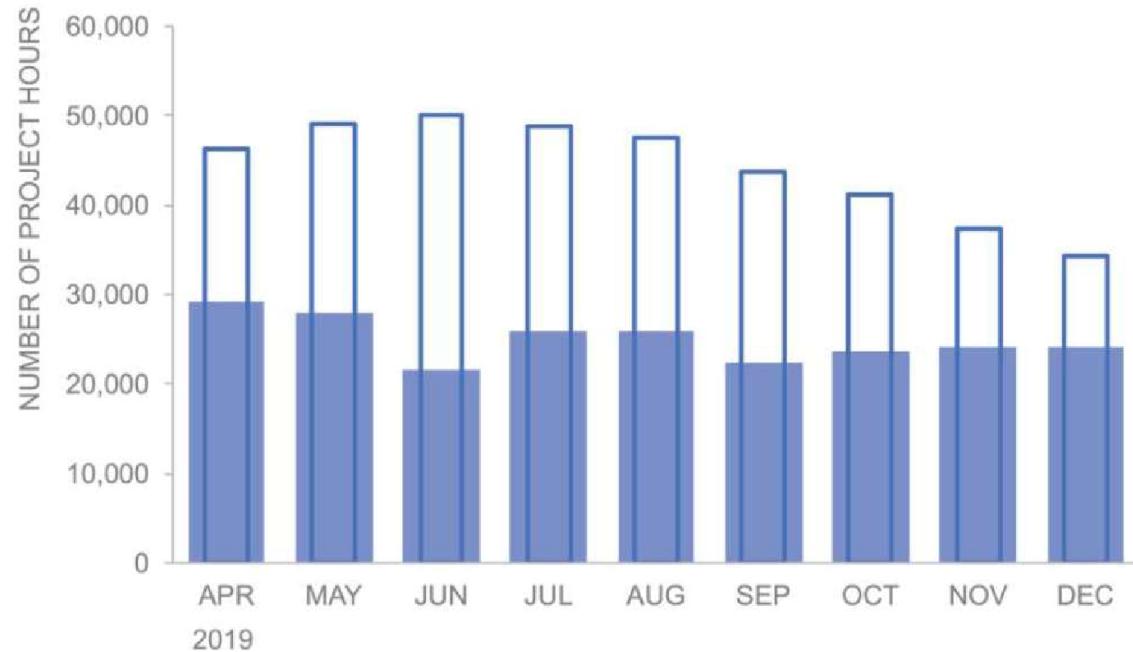


Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 5

Demand vs capacity over time

DEMAND | CAPACITY

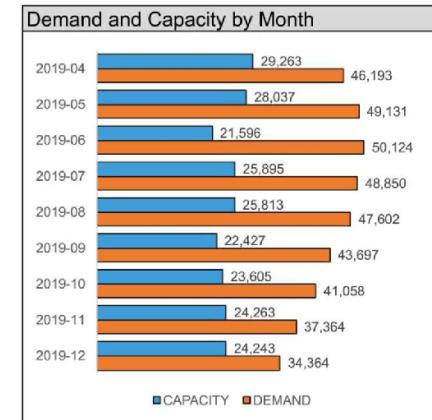
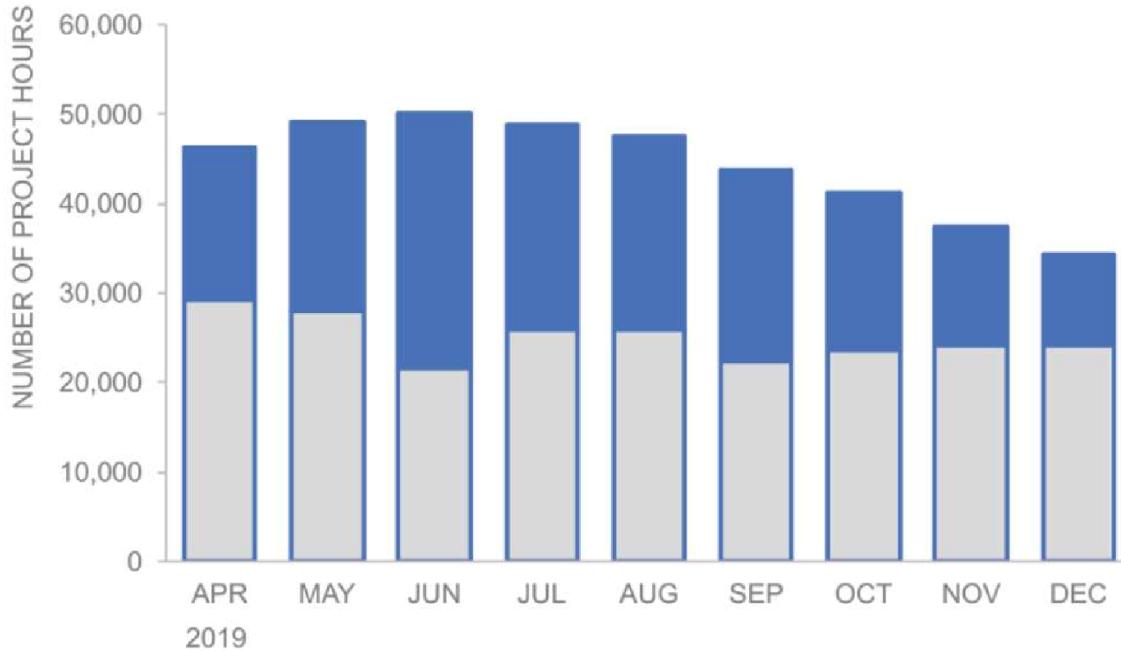


Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 5

Demand vs capacity over time

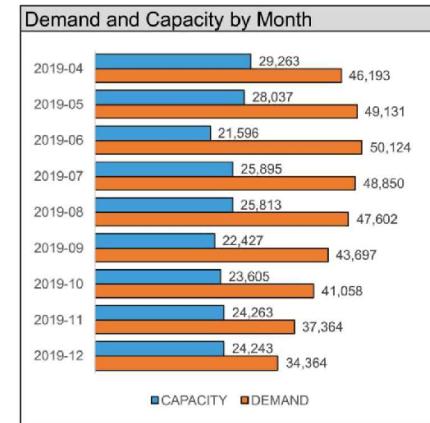
CAPACITY | UNMET DEMAND



Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 5

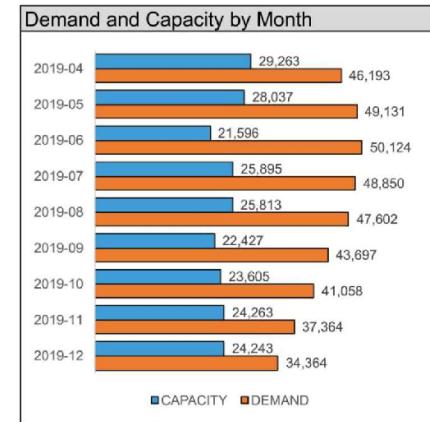
Demand vs capacity over time



Data visualisation : Fondamentaux de la data visualisation

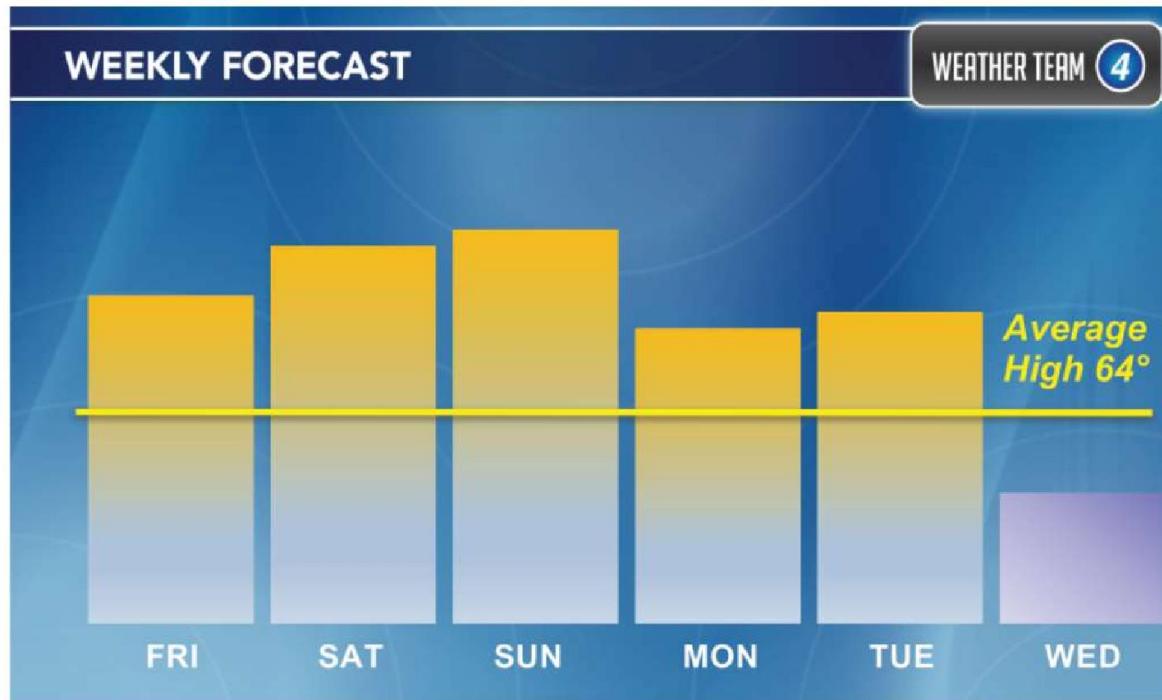
- **Comment choisir un graphique ? Exercice 5**

Unmet demand over time



Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 6
- Comment peut-on améliorer le visuel ci-dessous ?



Data visualisation : Fondamentaux de la data visualisation

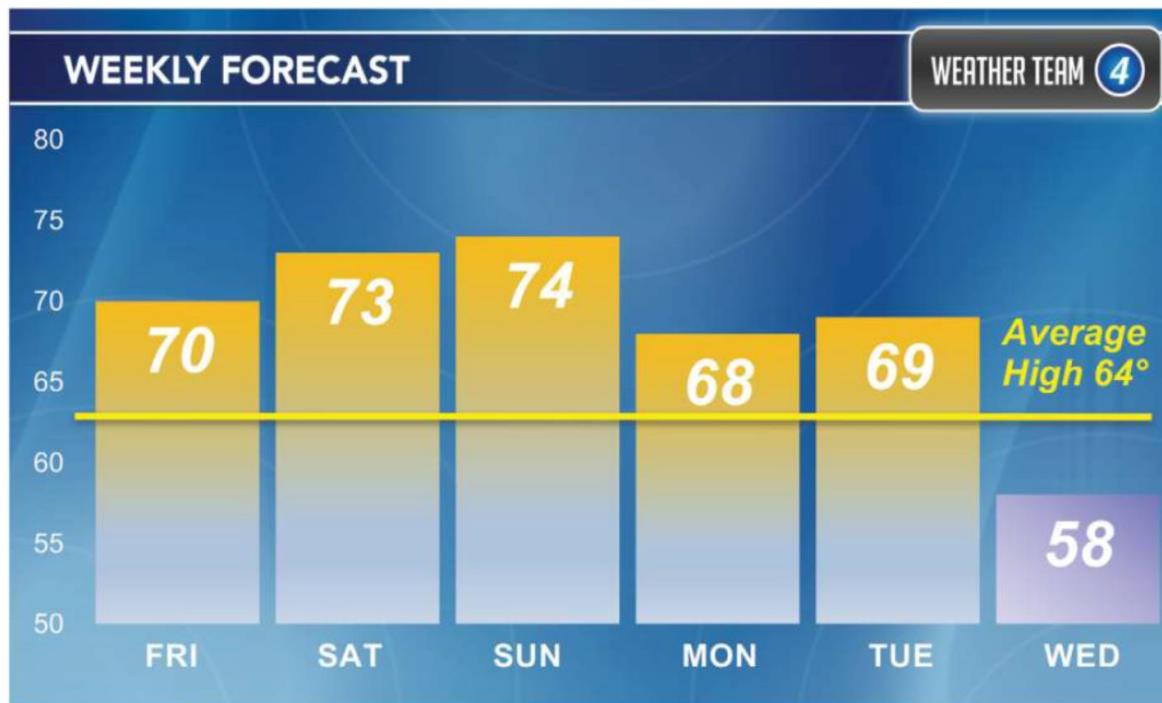
- Comment choisir un graphique ? Exercice 6
Quelle alternative est meilleure, et pourquoi ?



Data visualisation : Fondamentaux de la data visualisation

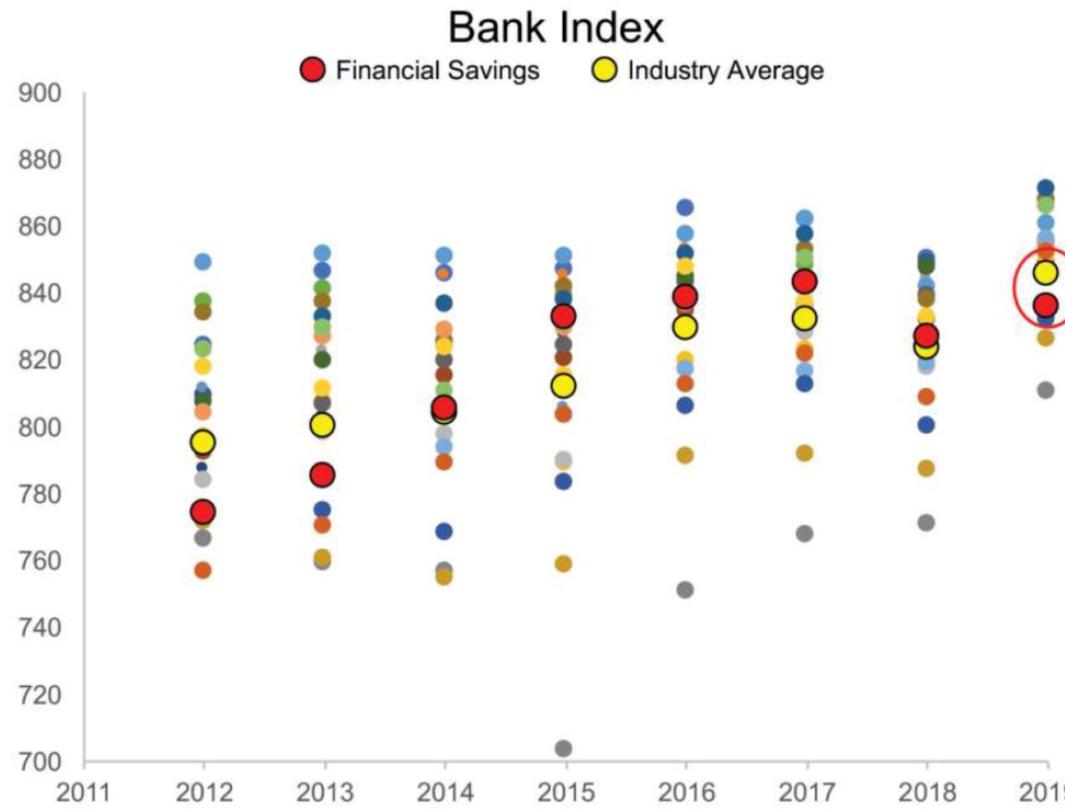
- Comment choisir un graphique ? Exercice 6

Quelle alternative est meilleure, et pourquoi ?



Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 7
- Comment peut-on améliorer le graphique ci-dessous ?

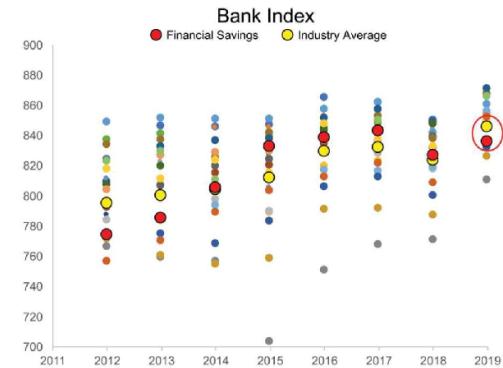
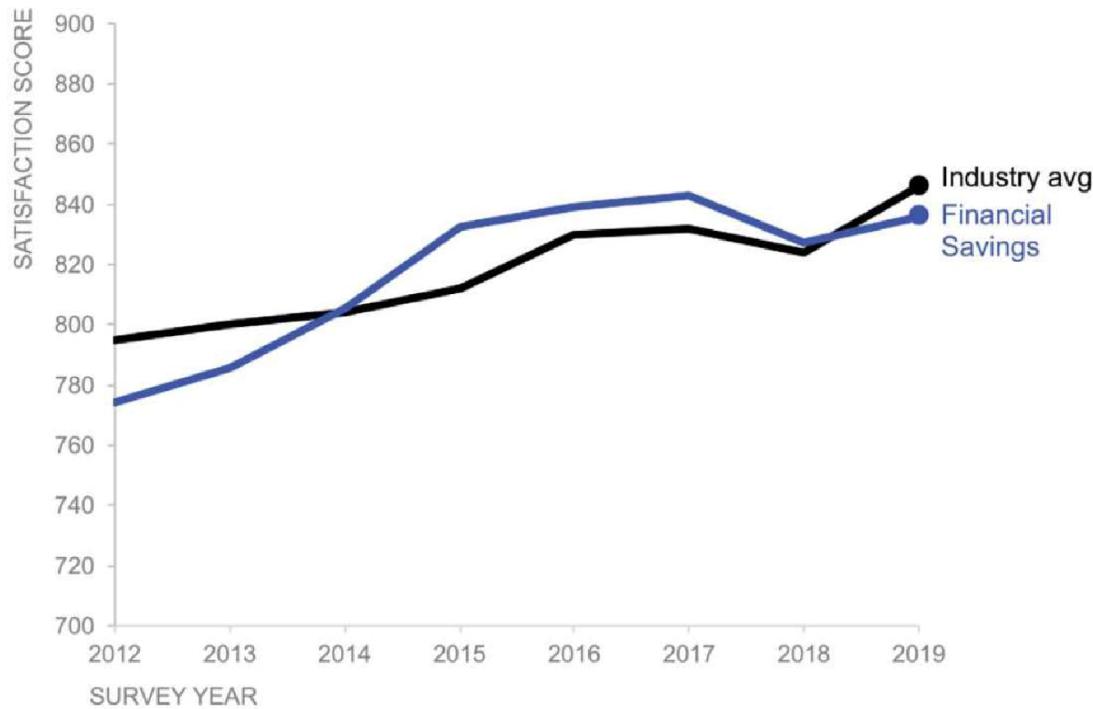


Data visualisation : Fondamentaux de la data visualisation

- Comment choisir un graphique ? Exercice 7

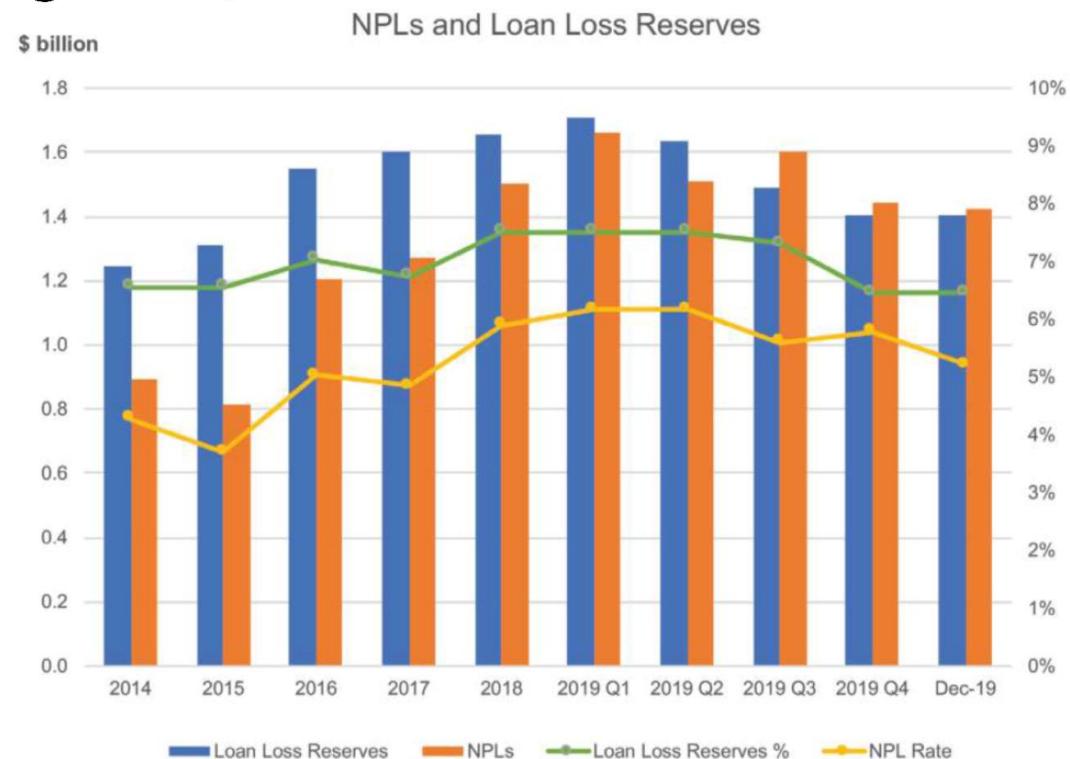
BRANCH SATISFACTION

Financial Savings below industry for first time in 5 years



Data visualisation : Fondamentaux de la data visualisation

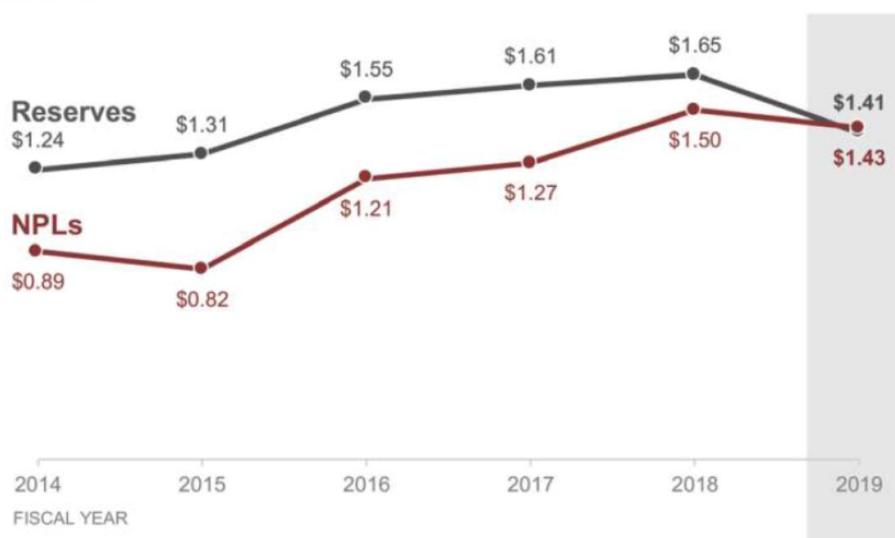
- Comment choisir un graphique ? Exercice 8
 - Comment peut-on améliorer le graphique ci-dessous ?
- (NLP = Non-Performing Loans)



Data visualisation : Fondamentaux de la data visualisation

• Comment choisir un graphique ? Exercice 8

Annual Loan Loss Reserves & Non-Performing Loans (NPLs)
BILLIONS



2019 quarterly view
BILLIONS

