

## Séance 4.2: Mesures de tendance centrale

### Discussion en classe

Visseho Adjiwanou, PhD.

30 January 2023

# A quoi ça sert?

- Une mesure de tendance centrale est une valeur **typique** ou **représentative** d'un ensemble de scores

## Résumé : Mesure de tendance centrale (paramètres de position)

Symbole	Définition	Formules
Moyenne	Somme des valeurs divisée par l'effectif de la série	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
Médiane	Valeur qui divise la distribution en deux parties égales	
Mode	Valeur observée de fréquence maximum	
Percentile	Valeurs qui divisent la distribution en 100 parties égales	

## Résumé : Mesure de dispersion

Symbole	Définition	Formules
Étendue	Différence entre la plus grande et la plus petite valeur de la variable	$G - P$
EIQ	3ème quartile - 1er quartile	$Q3 - Q1$
Déviations	La distance d'une valeur à la moyenne	$X - \bar{X}$
Sommes des carrés	Somme des carrés des déviations	$SC = \sum_{i=1}^n (X_i - \bar{X})^2$
Variance	Moyenne des carrés des déviations	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
Écart-type	Racine carrée de la variance	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$

## Résumé : Quel type de résumé pour quel type de variable?

Type de variable	Fréquence	Pourcentage	Commentaire
Nominale	Oui	Oui	Toujours
Ordinale	Oui	Oui	Toujours
Ratio/Intervalle	Pas souhaité	Pas souhaité	Oui si peu de modalités
Ratio/Intervalle (données groupées)	Oui	Oui	Toujours

# Résumé : Quel type de résumé pour quel type de variable?

Type de variable	Moyenne	Mode	Médiane	Variance	Écart-type
Nominale	Non	Oui	Non	Non	Non
Ordinale	Possible	Oui	Oui	Possible	Possible
Ratio/Intervalle	Oui	Oui	Oui	Oui	Oui
Ratio/Intervalle (données groupées)	Oui	Oui	Oui	Oui	Oui

## Exemple de calcul

## Distribution de revenu

Voici les revenus (en millier) d'un échantillon de 15 hommes et de 16 femmes

ind	Revenus	ind	Revenus	ind	Revenus	ind	Revenus
1	2	9	3.1	1	3.1	9	0.5
2	2.5	10	1.4	2	2.7	10	1.3
3	1.7	11	7.1	3	1.2	11	2.9
4	3	12	6.0	4	4.2	12	2.7
5	5	13	3.3	5	5.5	13	5.1
6	4.1	14	4.3	6	4.3	14	3.0
7	8.1	15	6.1	7	2.0	15	6.3
8	5.2			8	1.5	16	4.2



# Distribution du revenu des hommes

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
1.4	1	6.67	6.67	6.67	6.67
1.7	1	6.67	13.33	6.67	13.33
2	1	6.67	20.00	6.67	20.00
2.5	1	6.67	26.67	6.67	26.67
3	1	6.67	33.33	6.67	33.33
3.1	1	6.67	40.00	6.67	40.00
3.3	1	6.67	46.67	6.67	46.67
4.1	1	6.67	53.33	6.67	53.33
4.3	1	6.67	60.00	6.67	60.00
5	1	6.67	66.67	6.67	66.67
5.2	1	6.67	73.33	6.67	73.33
6	1	6.67	80.00	6.67	80.00
6.1	1	6.67	86.67	6.67	86.67
7.1	1	6.67	93.33	6.67	93.33
8.1	1	6.67	100.00	6.67	100.00
<NA>	0			0.00	100.00
Total	15	100.00	100.00	100.00	100.00

# Moyenne

- Revenu moyen des hommes

- $= (2 + 2.5 + 1.7 + 3 + 5 + 4.1 + 8.1 + 5.2 + 3.1 + 1.4 + 7.1 + 6.0 + 3.3 + 4.3 + 6.1) / 15$
- revenu moyen des hommes = 4.19

- Revenu moyen des femmes

- $= (3.1 + 2.7 + 1.2 + 4.2 + 5.5 + 4.3 + 2.0 + 1.5 + 0.5 + 1.3 + 2.9 + 2.7 + 5.1 + 3.0 + 6.3 + 4.2) / 16$
- Revenu moyen des femmes = 3.16

## Médiane du revenu des hommes

- **La Médiane** = valeur telle que la moitié des observations lui sont inférieures et donc la moitié lui sont supérieures.
- C'est donc assez facile à calculer, il suffit juste d'ordonner les cas.
- Milieu de la distribution =  $(15 + 1)/2 = 8$

# Médiane du revenu des hommes

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
1.4	1	6.67	6.67	6.67	6.67
1.7	1	6.67	13.33	6.67	13.33
2	1	6.67	20.00	6.67	20.00
2.5	1	6.67	26.67	6.67	26.67
3	1	6.67	33.33	6.67	33.33
3.1	1	6.67	40.00	6.67	40.00
3.3	1	6.67	46.67	6.67	46.67
4.1	1	6.67	53.33	6.67	53.33
4.3	1	6.67	60.00	6.67	60.00
5	1	6.67	66.67	6.67	66.67
5.2	1	6.67	73.33	6.67	73.33
6	1	6.67	80.00	6.67	80.00
6.1	1	6.67	86.67	6.67	86.67
7.1	1	6.67	93.33	6.67	93.33
8.1	1	6.67	100.00	6.67	100.00
<NA>	0			0.00	100.00
Total	15	100.00	100.00	100.00	100.00

##

## Médiane des femmes

- Milieu de la distribution :  $(16 + 1)/2 = 8.5$  entre la 8 et la 9e valeur

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0.5	1	6.25	6.25	6.25	6.25
1.2	1	6.25	12.50	6.25	12.50
1.3	1	6.25	18.75	6.25	18.75
1.5	1	6.25	25.00	6.25	25.00
2	1	6.25	31.25	6.25	31.25
2.7	2	12.50	43.75	12.50	43.75
2.9	1	6.25	50.00	6.25	50.00
3	1	6.25	56.25	6.25	56.25
3.1	1	6.25	62.50	6.25	62.50
4.2	2	12.50	75.00	12.50	75.00
4.3	1	6.25	81.25	6.25	81.25
5.1	1	6.25	87.50	6.25	87.50
5.5	1	6.25	93.75	6.25	93.75
6.6	1	6.25	100.00	6.25	100.00

## Médiane des femmes

- Médiane se situe entre la 8e et la 9e valeur
- Médiane =  $(2.9 + 3)/2 = 2.95$

# Mode

- Homme = il n'y a pas de mode
- Femme = deux modes, 2.7 et 4.2

## Premier quartile

Quelle est la localisation du premier quartile?

1	2	3	4	5	6	7	8	9	10	11	12	13	17
1.4	1.7	2	2.5	3	3.1	3.3	4.1	4.3	5	5.2	6	6.1	7.1



## Premier quartile

Quelle est la localisation du premier quartile?

$$L_k = k/100 * (n + 1)$$

k = 25 si premier quartile n = 15

donc  $l_{25} = 25/100 * (15+1) = 4$  Le premier quartile se trouve donc à la 4e position

premier quartile (Q1) = 2.5

## Troisième quartile

$$I_{75} = 75/100 * (15+1) = 12$$

$$\text{donc } Q3 = 6$$

## Et pour les femmes?

$$Q1 = ? \quad Q3 = ?$$

## Et pour les femmes

1	2		3	4		5	6	7	8		9	10	11	12	
.5	1	.2	1.3	1.5		2	2.7	2.7	2.9		3.	3.1	4.2	4.2	

## Et pour les femmes?

- $L_{25} = 25/100*(16+1) = 4.25$  (entre 4 et 5)

## Et pour les femmes?

- $L_{25} = 25/100*(16+1) = 4.25$  (entre 4 et 5)
- $L_{75} = 75/100*(16+1) = 12.75$  (entre 12 et 13)

## Et pour les femmes?

- $L_{25} = 25/100*(16+1) = 4.25$  (entre 4 et 5)
- $L_{75} = 75/100*(16+1) = 12.75$  (entre 12 et 13)
- $Q1 = (1.5 + 2)/2 = 1.75$

## Et pour les femmes?

- $L_{25} = 25/100*(16+1) = 4.25$  (entre 4 et 5)
- $L_{75} = 75/100*(16+1) = 12.75$  (entre 12 et 13)
- $Q1 = (1.5 + 2)/2 = 1.75$
- $Q3 = (4.2 + 4.3)/2 = 4.25$

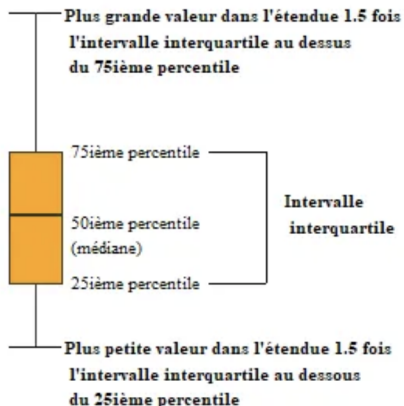


## Et pour les femmes?

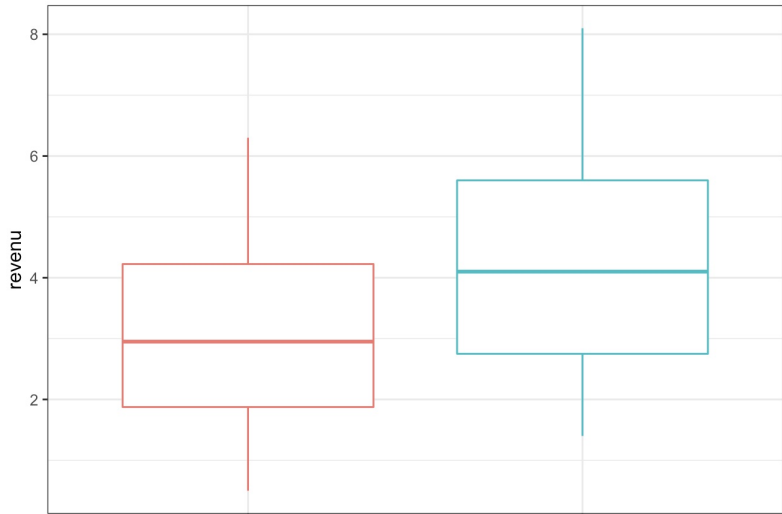
- $L_{25} = 25/100*(16+1) = 4.25$  (entre 4 et 5)
- $L_{75} = 75/100*(16+1) = 12.75$  (entre 12 et 13)
- $Q1 = (1.5 + 2)/2 = 1.75$
- $Q3 = (4.2 + 4.3)/2 = 4.25$
- Une difficulté arrive avec les variables ordinales

# Représentation

- Boxplot



# Représentation



# Médiane des variables ordinales

Exemple: Attitude envers les immigrants et les emplois

- Q1: Sur une échelle de 1 (totalement en désaccord) à 5 (totalement d'accord), que pensez-vous de l'affirmation suivante: "Les immigrants volent nos emplois"

## Médiane des variables ordinales

Valeur	Fréquence	Fréq. cumulée	Pourcentage	Pourc. cumulé
1	170	170		
2	446	616		
3	299	915		
4	301	1216		
5	65	1281		
N	1281			

- Médiane va diviser la distribution en deux parties égales

## Médiane des variables ordinales

Valeur	Fréquence	Fréq. cumulée	Pourcentage	Pourc. cumulé
1	170	170		
2	446	616		
3	299	915		
4	301	1216		
5	65	1281		
N	1281			

- Médiane va diviser la distribution en deux parties égales
- $(1281 + 1)/2 = 641$

## Médiane des variables ordinales

Valeur	Fréquence	Fréq. cumulée	Pourcentage	Pourc. cumulé
1	170	170		
2	446	616		
3	299	915		
4	301	1216		
5	65	1281		
N	1281			

- Médiane va diviser la distribution en deux parties égales
- $(1281 + 1)/2 = 641$
- la médiane est la valeur du 641<sup>e</sup> score, c'est-à-dire quelque part parmi les 299 scores 3

## Médiane des variables ordinales

Valeur	Fréquence	Fréq. cumulée	Pourcentage	Pourc. cumulé
1	170	170		
2	446	616		
3	299	915		
4	301	1216		
5	65	1281		
N	1281			

- Médiane va diviser la distribution en deux parties égales
- $(1281 + 1)/2 = 641$
- la médiane est la valeur du 641<sup>e</sup> score, c'est-à-dire quelque part parmi les 299 scores 3



## Médiane des variables ordinales

- Convention: on va supposer que derrière les réponses à cette question, il y a une échelle (ratio) allant de 0.5 à 5.5.

## Médiane des variables ordinales

- Convention: on va supposer que derrière les réponses à cette question, il y a une échelle (ratio) allant de 0.5 à 5.5.
- Autrement dit, ceux qui ont répondu 1, aurait répondu en réalité entre 0.5 et 1.5. On les ramène donc à la moyenne de l'intervalle qui vaut  $(0.5 + 1.5)/2 = 1$

## Médiane des variables ordinales

- Convention: on va supposer que derrière les réponses à cette question, il y a une échelle (ratio) allant de 0.5 à 5.5.
- Autrement dit, ceux qui ont répondu 1, aurait répondu en réalité entre 0.5 et 1.5. On les ramène donc à la moyenne de l'intervalle qui vaut  $(0.5 + 1.5)/2 = 1$
- Ainsi, les 299 personnes qui ont répondu 3 ont en fait répondu entre 2.5 et 3.5

## Médiane des variables ordinales

- Convention: on va supposer que derrière les réponses à cette question, il y a une échelle (ratio) allant de 0.5 à 5.5.
- Autrement dit, ceux qui ont répondu 1, aurait répondu en réalité entre 0.5 et 1.5. On les ramène donc à la moyenne de l'intervalle qui vaut  $(0.5 + 1.5)/2 = 1$
- Ainsi, les 299 personnes qui ont répondu 3 ont en fait répondu entre 2.5 et 3.5
- Nous allons donc interpoler pour trouver à quel endroit se situe la médiane

## Médiane des variables ordinales

$$617 \text{ ————— } 641 \text{ ————— } 617 + 299 = 916$$

$$2.5 \text{ ————— } M \text{ ————— } 3.5$$

# Médiane des variables ordinales

- Interpolation

# Médiane des variables ordinales

- Interpolation
- Application de la règle de trois

# Médiane des variables ordinales

- Interpolation
- Application de la règle de trois
- Si 3 pains coûtent 55\$, combien coûte 2 pains?



## Médiane des variables ordinales

$$617 \text{ ————— } 641 \text{ ————— } 617 + 299 = 916$$

$$2.5 \text{ ————— } M \text{ ————— } 3.5$$

$$(M - 3.5)/(2.5 - 3.5) = (641 - 916)/(617 - 916)$$

$$M = [(641 - 916)/(617 - 916)]/(2.5 - 3.5) + 3.5$$

$$M = 2.6$$

## Vous pouvez préféré utiliser la formule du livre

$$Md = L + \left( \frac{N/2 - F}{f} \right)(i)$$

- L = la limite inférieure de l'intervalle contenant la médiane (2.5)
- N = le nombre de cas (1281)
- F = la fréquence cumulative des scores inférieurs à l'intervalle contenant la médiane (616)
- f = le nombre de scores que comprend l'intervalle contenant la médiane (299)
- i = la largeur de l'intervalle contenant la médiane (1)
- $Md = 2.5 + (1281/2 - 616)/299 * 1 = 2.6$

## Exemple d'application : mesurer l'inégalité

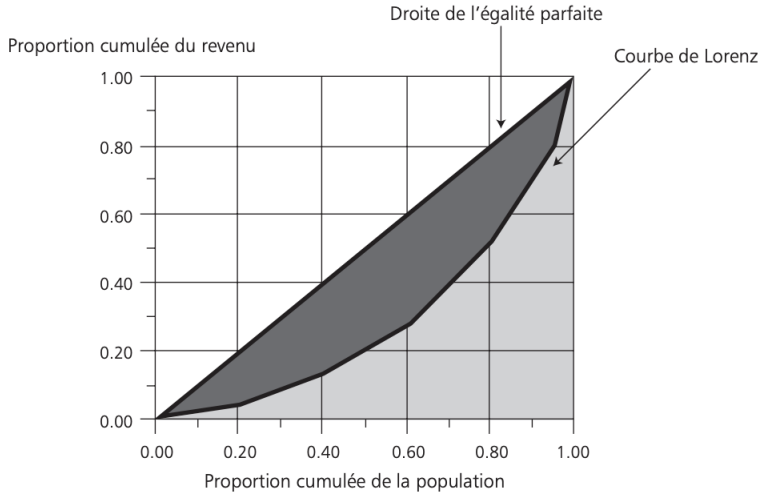
# Mesure de l'inégalité

- Moyenne
- Médiane
- Coefficient de Gini
- Indice de Palma

## Coefficient de Gini

- Le coefficient de Gini sert à mesurer le niveau d'inégalité de la répartition du revenu au sein de diverses populations ou au sein de la même population au fil du temps.
- Il peut être calculé pour d'autres indicateurs autre que le revenu.
- La courbe de Lorenz (figure) présente le rapport entre la proportion cumulée de la population, ordonnée selon le niveau de revenu, et la proportion cumulée du revenu total lui revenant.
- Le coefficient de Gini provient du calcul de l'aire entre la courbe de Lorenz et la droite de l'égalité parfaite.

# Coefficient de Gini



## Coefficient de Gini

- Plus cette aire est grande, plus l'inégalité est marquée. Les valeurs du coefficient de Gini varient entre 0 et 1.
- Une valeur de 0 indique que le revenu est également divisé entre tous les membres de la population, ceux-ci recevant exactement la même somme de revenu.

## Coefficient de Gini

- Plus cette aire est grande, plus l'inégalité est marquée. Les valeurs du coefficient de Gini varient entre 0 et 1.
- Une valeur de 0 indique que le revenu est également divisé entre tous les membres de la population, ceux-ci recevant exactement la même somme de revenu.
- Par contre, une valeur de 1 dénote une distribution parfaitement inégale au sein de laquelle une seule unité possède l'ensemble du revenu de l'économie.




# Coefficient de Gini

- Plus cette aire est grande, plus l'inégalité est marquée. Les valeurs du coefficient de Gini varient entre 0 et 1.
- Une valeur de 0 indique que le revenu est également divisé entre tous les membres de la population, ceux-ci recevant exactement la même somme de revenu.
- Par contre, une valeur de 1 dénote une distribution parfaitement inégale au sein de laquelle une seule unité possède l'ensemble du revenu de l'économie.
- Une diminution de la valeur du coefficient de Gini peut être interprétée comme une diminution de l'inégalité, et vice versa.

# Indice de Palma

<https://jeanneemard.wordpress.com/2022/05/19/les-inegalites-au-quebec-de-1976-a-2020-selon-lindice-de-palma-et-le-coefficient-de-gini/>

10%	Décile 1	40% revenus les plus bas		
10%	Décile 2			
10%	Décile 3			
10%	Décile 4			
10%	Décile 5	<i>Groupe généralement désigné comme «classe moyenne»</i>		
10%	Décile 6			
10%	Décile 7			
10%	Décile 8			
10%	Décile 9			
10%	Décile 10	10% revenus les plus élevés		

Palma =  
ratio entre  
les revenus  
détenus  
par ces  
deux  
groupes

# Indice de Palma

- IP est obtenu en divisant le revenu total des membres du 10e décile par le revenu total des membres des quatre premiers déciles (les plus pauvres)
- Un IP égal à 1 signifie que les ménages du décile le plus riche gagnent en moyenne quatre fois plus que ceux des quatre déciles les plus pauvres,
- Un IP égal à 2 équivaut à huit fois plus et ainsi de suite.

## Quel revenu utilisé?

- Le **revenu ajusté** est obtenu en divisant le revenu total de tous les membres d'un ménage par la racine carrée de la taille de ce ménage.
- Par exemple, le revenu ajusté des membres d'un ménage de quatre personnes qui a un revenu total de 100 000\$ sera de 50 000 \$ ( $100\,000 \$ / \sqrt{4}$ , soit 2, = 50 000 \$) et ce revenu sera accordé aux quatre membres de ce ménage.

## Quel revenu utilisé?

- Le CG et l'IP peuvent être présentés en fonction de trois types de revenus :
  - selon le **revenu du marché** : somme des revenus d'emploi (travail salarié ou montant net de travail autonome), de placements, de retraite (régime privé de pension) et autres;

## Quel revenu utilisé?

- Le CG et l'IP peuvent être présentés en fonction de trois types de revenus :
  - selon le **revenu du marché** : somme des revenus d'emploi (travail salarié ou montant net de travail autonome), de placements, de retraite (régime privé de pension) et autres;
  - selon le **revenu total** : revenu du marché plus les transferts gouvernementaux (aide sociale, assurance-emploi, pensions de la sécurité de la vieillesse, supplément de revenu garanti, prestations du Régime des rentes du Québec, etc.), avant impôt;

## Quel revenu utilisé?

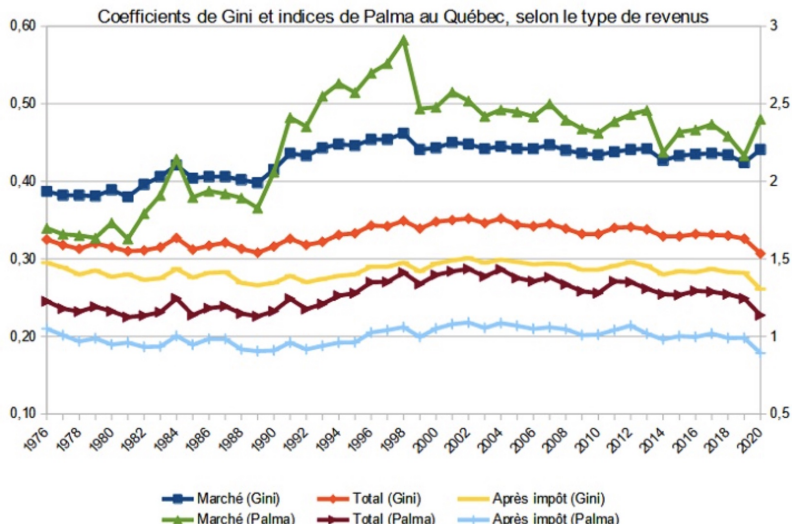
- Le CG et l'IP peuvent être présentés en fonction de trois types de revenus :
  - selon le **revenu du marché** : somme des revenus d'emploi (travail salarié ou montant net de travail autonome), de placements, de retraite (régime privé de pension) et autres;
  - selon le **revenu total** : revenu du marché plus les transferts gouvernementaux (aide sociale, assurance-emploi, pensions de la sécurité de la vieillesse, supplément de revenu garanti, prestations du Régime des rentes du Québec, etc.), avant impôt;
  - selon le **revenu après impôt** : revenu total moins l'impôt sur le revenu.

## Quel revenu utilisé?

- Le CG et l'IP peuvent être présentés en fonction de trois types de revenus :
  - selon le **revenu du marché** : somme des revenus d'emploi (travail salarié ou montant net de travail autonome), de placements, de retraite (régime privé de pension) et autres;
  - selon le **revenu total** : revenu du marché plus les transferts gouvernementaux (aide sociale, assurance-emploi, pensions de la sécurité de la vieillesse, supplément de revenu garanti, prestations du Régime des rentes du Québec, etc.), avant impôt;
  - selon le **revenu après impôt** : revenu total moins l'impôt sur le revenu.
  - Le CG et l'IP selon le revenu après impôt sont les plus couramment utilisés, car ils représentent les inégalités après intervention gouvernementale (transferts et impôt) et donc



# Évolution de l'inégalité au Québec



Visseho Adjwanou, PhD.

Séance 4.2: Mesures de tendance centrale

# Pour la semaine prochaine

## 1 Lecture

- Paramètres de variation (ou de dispersion) - Fox : chapitre 4, pp.91-103
- Distribution d'échantillonnage - Fox : Chapitre 4, pp.103-120

## 2 Application

- <https://juba.github.io/tidyverse/01-presentation.html>
- [https://juba.github.io/tidyverse/02-prise\\_en\\_main.html](https://juba.github.io/tidyverse/02-prise_en_main.html)
- [https://juba.github.io/tidyverse/03-premier\\_travail.html](https://juba.github.io/tidyverse/03-premier_travail.html)

```
ggplot(donnee_revenu) +  
  geom_boxplot(aes(x = sexe, y = revenu, color = sexe)) +  
  theme_bw()
```

