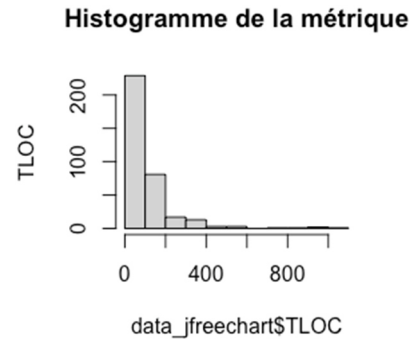
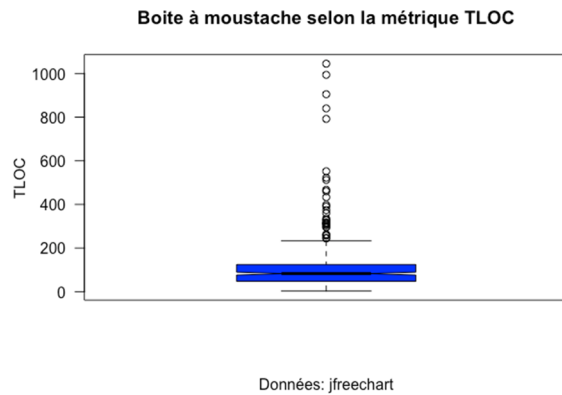


## Tâche 1 : Visualisation des métriques de l'échantillon.

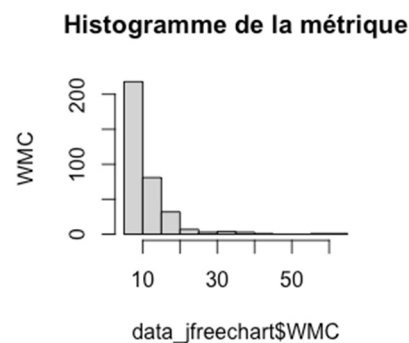
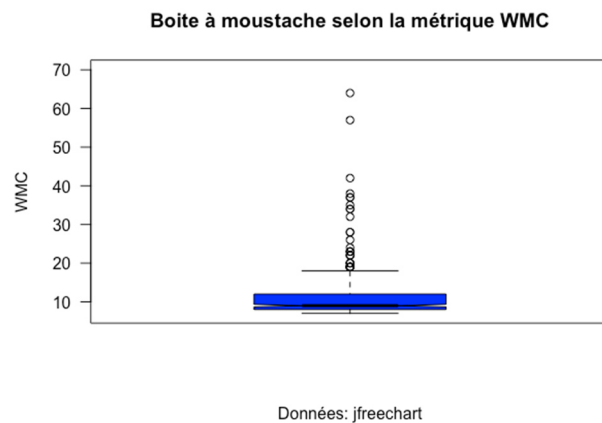
Boîtes à moustache :

### 1- TLOC



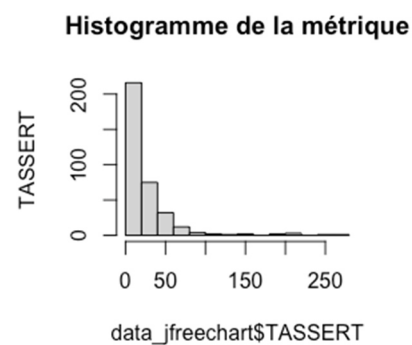
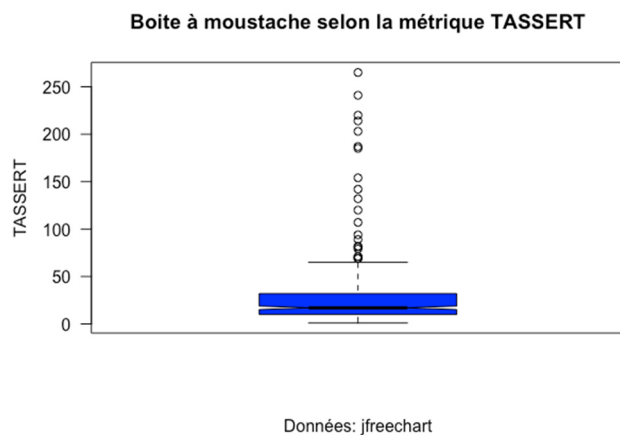
Min	1st Quartile	Median	Mean	3rd Quartile	Max
3	47.5	83.0	115.1	124.5	1045

### 2- WMC



Min	1st Quartile	Median	Mean	3rd Quartile	Max
7	8	9	11,58	12	64

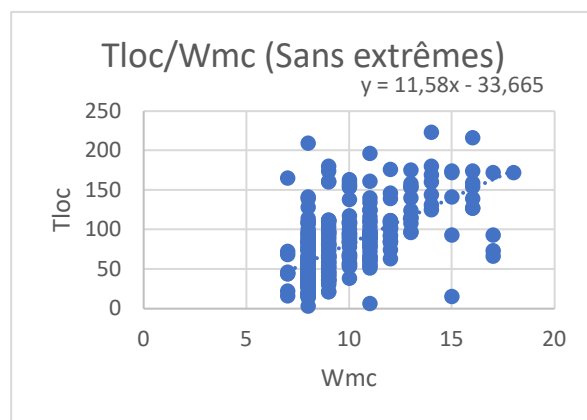
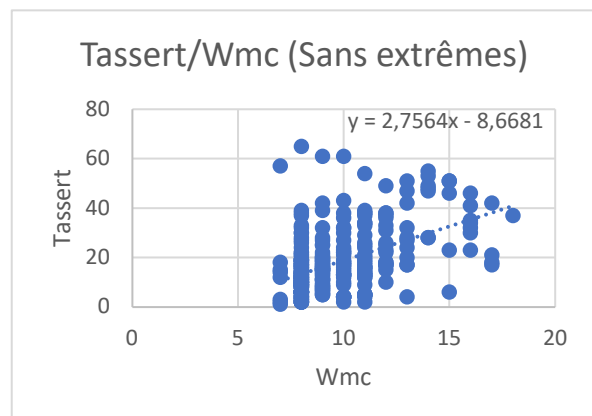
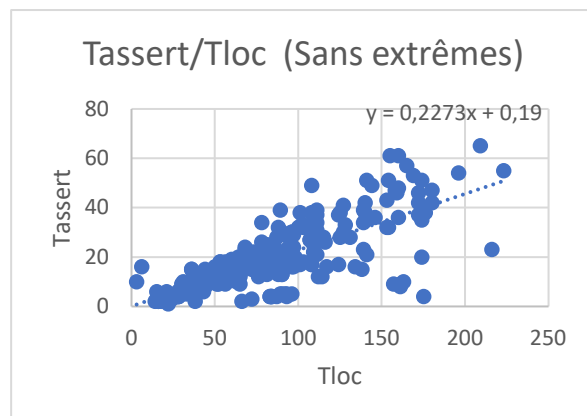
### 3-TASSERT



Min	1st Quartile	Median	Mean	3rd Quartile	Max
1	10	17	27.19	32	265

Nous observons que les histogrammes de Tloc, Tassert et Wmc sont tous les 3 similaires et semblent suivre une courbe de forme  $\frac{1}{a \cdot x} + b$ . De plus, la plupart des points extrêmes trouvés par une boîte à moustache se retrouvent aussi dans les points extrêmes des autres boîtes à moustaches. Il semble donc y avoir un lien entre les métriques Tloc, Tassert et Wmc.

## Tâche 2 : Étudier les corrélations des métriques.



Coefficient			
Corrélation	Tassert/TLOC	Tassert/WMC	Tloc/Wmc
Avec extrêmes	0,938750574	0,768556631	0,783622162
Sans extrêmes	0,774951952	0,48204681	0,594096225

Pour étudier la corrélation entre Tassert et Tloc et Wmc, nous avons fait des graphiques à nuage de points, puis calculé les droites de régression correspondantes. Par curiosité, nous l'avons aussi fait entre Tloc et Wmc pour avoir une idée de la corrélation entre les deux. Nous savons que Tloc et Wmc sont deux bonnes métriques permettant de mesurer la complexité d'un programme ce qui nous sera utile lorsque nous voudrions évaluer le coefficient de corrélation obtenu entre Tassert et Tloc et Wmc. Nous avons ensuite utilisé les boîtes à moustaches pour refaire ces mêmes graphiques en ayant enlevé les mesures extrêmes. Après réflexion, nous avons décidé d'utiliser les graphes sans mesures extrêmes, car celles-ci avaient un impact trop important sur la courbe de tendance.

Après mesure des coefficients, nous obtenons  $r = 0,77$  entre Tassert et Tloc et  $r = 0,48$  entre Tassert et Wmc. Nous avons donc une corrélation forte entre Tassert et Tloc et une faible entre Tassert et Wmc. Nous observons aussi que la corrélation entre Tloc et Wmc est faible ( $r = 0,59$ ), ce qui semble indiquer

ce que Tassert mesure est similaire à Tloc, ce qui est logique puisque Tassert mesure le nombre de tests dans une classe et Tloc mesure le nombre de lignes de code d'une classe. En pratique, il sera très difficile d'avoir beaucoup de tests dans une classe tout en ayant un nombre limité de lignes de code. À l'inverse, Wmc mesure la somme des complexités des méthodes dans une classe. Ainsi, une classe possédant un grand Tassert pourrait bien n'être composée que de méthodes simples. Il est donc logique que la corrélation entre Tloc et Tassert soit plus forte qu'entre Wmc et Tassert.

### Tâche 3 : Évaluation de l'hypothèse.

**Hypothèse à évaluer :** « les classes qui contiennent plus de 20 assertions sont plus complexes que celles qui contiennent moins de 20 assertions »

**Groupe de comparaison :** Divisions des classes en deux groupes, celles avec plus de 20 assertions et celles avec un Tassert inférieur ou égal à 20 assertions.

#### Hypothèses:

- **Hypothèse nulle (H0) :** Il n'y a pas de différence significative de complexité entre les classes avec plus de 20 assertions et celles avec moins de 20 assertions.
- **Hypothèse alternative (H1) :** Les classes avec plus de 20 assertions sont plus complexes que celles avec moins de 20 assertions.

#### Définition des variables:

- **Variable indépendante :** Le nombre d'assertions dans une classe
- **Variable dépendante :** La complexité de la classe (mesurée par TLOC, WMC)

#### Interprétation et généralisation des résultats :

```
> summary(classes_moins_egale_a_20_assertions$TLOC)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  3.00  37.50  55.50  60.52  80.25 175.00
> summary(classes_plus_de_20_assertions$TLOC)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  68.0  105.0  141.0  202.5  221.5 1045.0
> summary(classes_moins_egale_a_20_assertions$WMC)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  7.000  8.000  8.000  9.139 10.000 20.000
> summary(classes_plus_de_20_assertions$WMC)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  7.0  10.0  13.0  15.5  18.0  64.0
```

Dans les statistiques obtenues, pour chaque type de classe (celles avec des TASSERT ≤ 20 et celles avec TASSERT > 20) nous remarquons que les classes avec un TASSERT > 20 ont tendance à avoir plus de lignes, car le TLOC minimum pour ces classes est 68 tandis que celle avec moins de 20 assertions ont un minimum de 3 lignes de code qui ne sont pas des commentaires.

La médiane de TLOC est de 55.5, ce qui signifie que 50% des classes ont 55.5 lignes de code ou moins pour les classes avec un max de 20 assertions, celles pour les classes avec plus de 20 assertions est 141 ce qui indique que 50% de ces classes ont 141 lignes de code ou moins.

Dans de l'analyse à partir de la métrique TLOC, les classes avec plus d'assertions ont une variation plus importante dans les tailles des classes. Les classes avec plus de 20 assertions ont tendance à être plus grandes en termes de nombre de lignes de code, avec une moyenne et une médiane significativement plus élevées que celles avec moins de 20 assertions.

Nous avons aussi décidé d'aller plus loin dans l'analyse en utilisant la métrique WMC qui permet d'évaluer la complexité d'une classe dans le contexte de la programmation orientée objet.

La médiane de WMC est de 8 pour les classes avec un maximum de 20 assertions, ce qui indique que 50% des classes ont un WMC de 8 ou moins. Elle est de 13 pour celles avec plus de 20 assertions. La moyenne est de 15.5 pour les classes avec plus de 20 assertions, cela montre une tendance centrale vers des valeurs plus élevées. Les classes avec plus de 20 assertions ont tendance à avoir un WMC plus élevé, indiquant une complexité plus importante par rapport aux classes avec moins de 20 assertions.

Les statistiques descriptives montrent que les classes avec plus de 20 assertions sont généralement plus grandes (en termes de lignes de code) et plus complexes (mesurées par WMC) que celles avec moins de 20 assertions. Cela peut partiellement confirmer que les classes avec plus d'assertions sont plus complexes. IL est cependant important de ne pas se limiter à cette étude pour affirmer que les classes de plus de 20 assertions sont plus complexes, car plusieurs autres facteurs internes et externes peuvent entrer en jeu.

#### Discussion des menaces à la validité :

S'assurer que la sélection des groupes est représentative de la population, certaines variables peuvent influencer à la fois le nombre d'assertions et la complexité des classes, la généralisation de notre étude peut être limitée, car l'échantillonnage dans le cas où l'échantillonnage n'est pas représentatif. Les mesures utilisées pour évaluer la complexité ne sont pas suffisantes.

