

Mais qui a des téléphones ?!

Justine Blanchot et Elyass Sayd, Groupe 0

25/04/2022

Introduction

Avant la fameuse conférence *Apple Keynote* de 2007 par Steve Jobs où il annonçait la sortie du premier iPhone, personne n'aurait pu imaginer que les écrans tactiles se retrouveraient dans toutes les mains. L'iPhone était le précurseur du marché des smartphones qui a connu une des expansions les plus fulgurantes avec l'amélioration constante des infrastructures internet. Aurions-nous pu prévoir quels marchés allaient-être les plus réceptifs à l'arrivée de cette technologie? À l'aide de quels critères? Nous allons au fil de ce rapport nous questionner sur les corrélations possibles pour un pays donné entre le nombre de téléphones pour mille habitants et différents critères de développement.

Problématique et objectif

A partir d'un jeu de données issu de Kaggle, publié en 2017 par Fernando Lasso, intitulé "Countries of the world" et rassemblant des données du gouvernement américain (Anciennes données du CIA World Factbook et International Consortium for the Advancement of Academic Publication <https://gsociology.icaap.org/datasetupload.html>), nous allons essayer d'expliquer la déjà existante fougue du téléphone mobile en 2000 en fonction de certains indicateurs.

Nous tenterons donc de répondre à la problématique suivante:

Aurions-nous pu anticiper la réussite du marché des téléphones portables dans un pays à l'aide des indicateurs classiques de développement? Si oui, à l'aide desquels? Si le temps nous le permet, nous comparerons notre analyse avec nouvelles données plus récentes du World Factbook.

I - Analyse descriptive des données

A - Analyse du jeu de données Le jeu de données que nous avons choisi contient les données de 227 pays par rapport à 19 variables. Dans le cadre de cette étude, nous nous limiterons aux variables suivantes: la région d'appartenance du pays, la densité de population (nombre habitants par miles carré), le PIB par habitant, le taux de natalité et, enfin, la part de la population active travaillant dans l'agriculture, l'industrie et les services.

```
Registered S3 method overwritten by 'GGally':
```

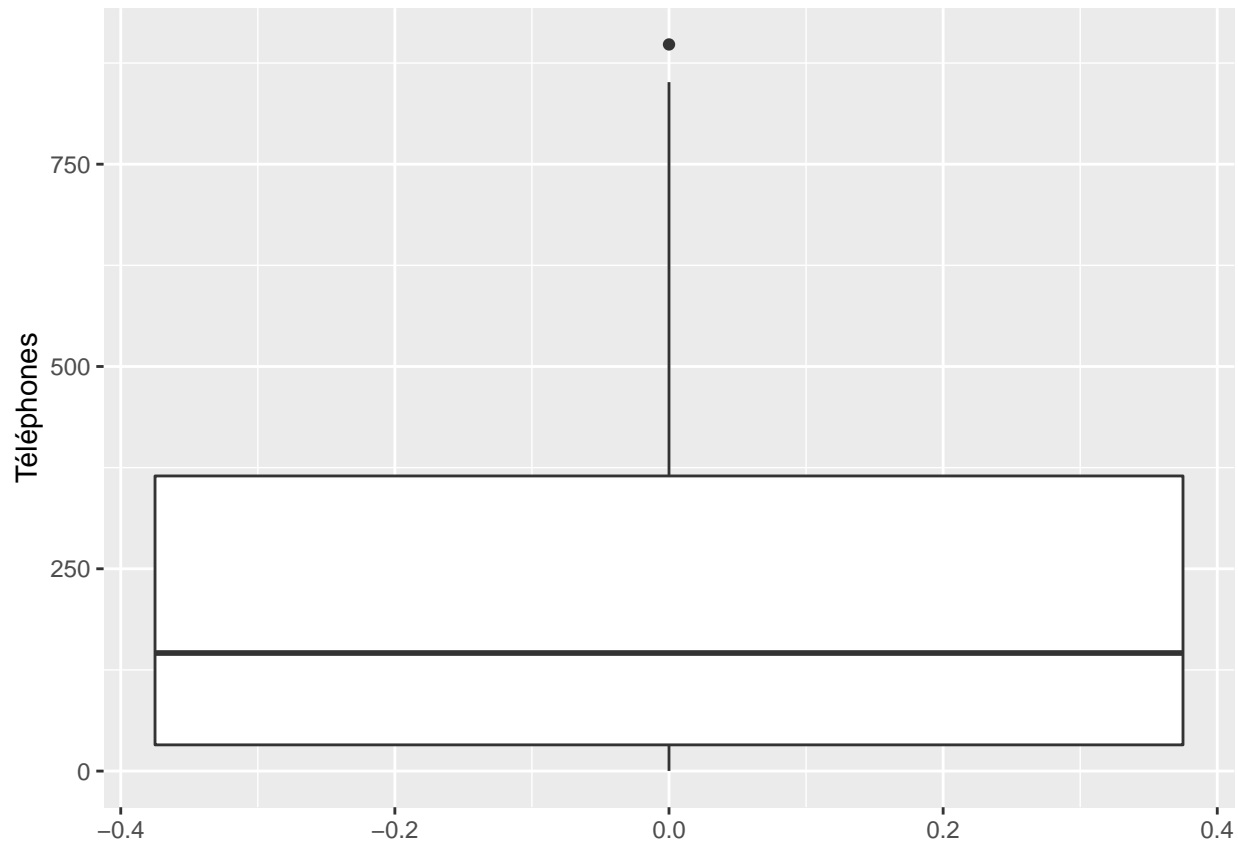
```
  method from  
+ .gg      ggplot2
```

Le chargement a nécessité le package : carData

Pour être en adéquation avec les consignes du projet, nous avons choisi 7 variables d'intérêt, puisqu'il était recommandé d'en prendre 3 ou 4 minimum. La région d'appartenance du pays est une variable qualitative. Nos autres variables sont quantitatives. Les noms des colonnes du fichier ont été renommées en français. Notre variable d'intérêt possédait 4 valeurs manquantes. C'est également le cas pour certaines de nos variables explicatives. Nous avons décidé de retirer les pays où il manque des données de notre étude. Il reste donc 208, 9 pays. Il était demandé d'avoir au minimum 100 données, ce qui est donc le cas dans notre étude.

B - Analyse univariée de la variable d'intérêt Introduisons des notations: T , pour téléphones, est la variables à expliquer, c'est-à-dire le nombres de téléphones pour mille habitants. T est la réalisation d'une variable aléatoire X . Dans cette première partie, nous allons analyser les données de T et essayer de déterminer la loi de X . Commençons par regarder quelques caractéristiques de T :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2	32.4	145.9	222.3	364.7	898.0



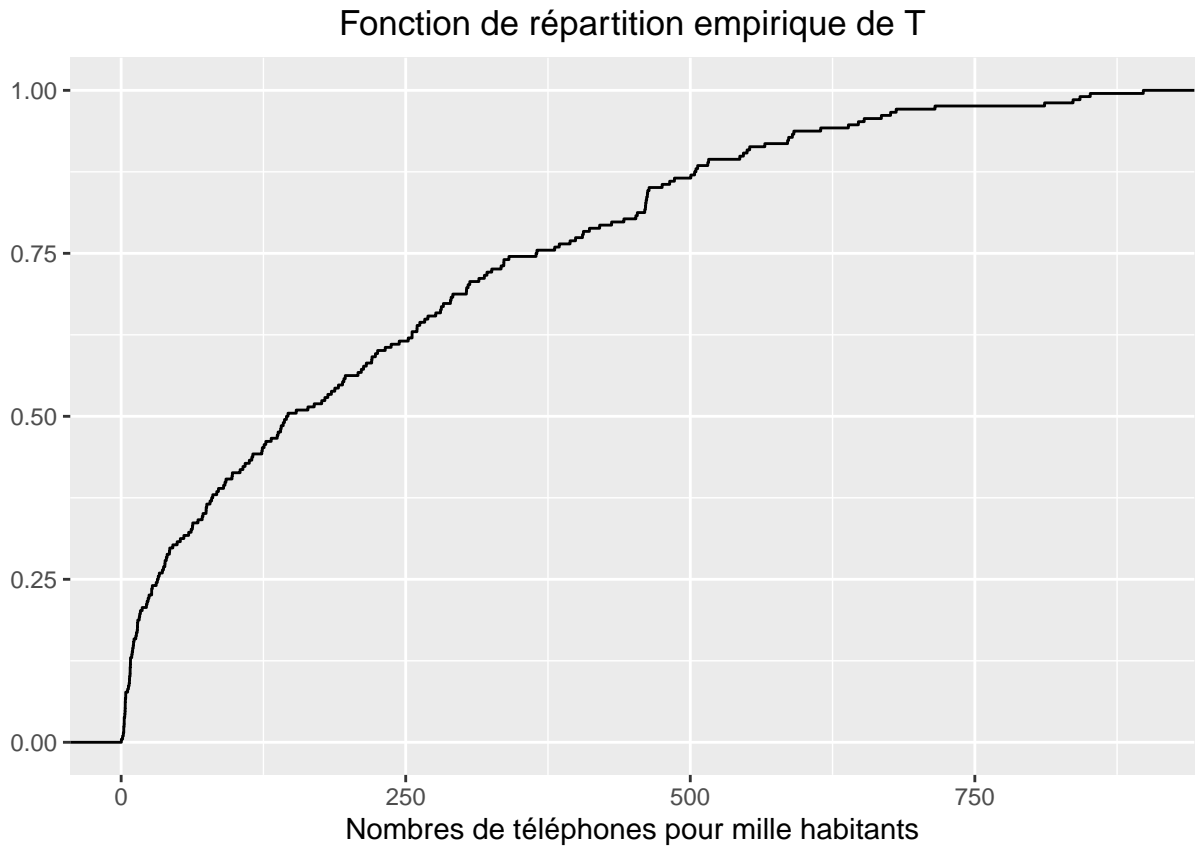
On peut commencer à analyser les indicateurs statistiques de notre échantillon. Par lecture de notre sommaire et de notre box-plot, on obtient que la loi n'est pas symétrique, puisque la moyenne, 222.2533654 est différente de la médiane, 145.9.

Les deux indicateurs de tendance centrale sont assez bas. Le premier quartile est à 32.4 mais surtout le troisième quartile est égal à 32.4. Cela signifie que 75% des pays ont moins de 364.7 téléphones pour mille habitants.

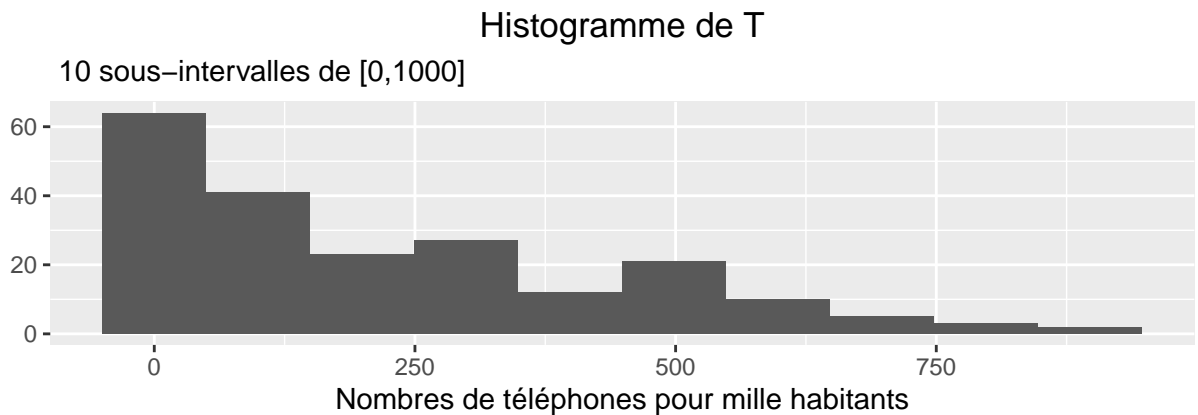
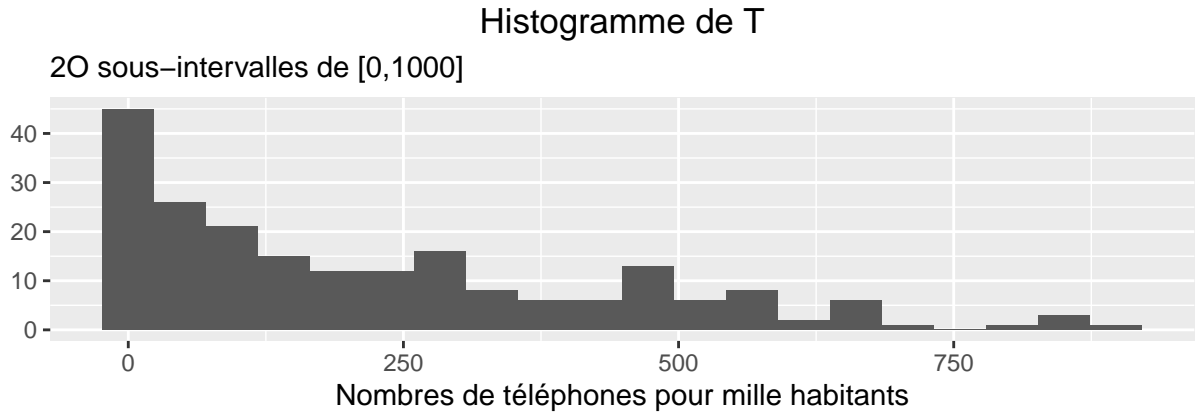
L'étendue est de 897.8, mais elle est sensible aux valeurs aberrantes et nous voyons dans notre box-plot que nous en avons une. De plus, la longueur d'une des moustaches est assez élevée. L'écart-interquartile, plus robuste que l'étendue, est de 332.3. Ainsi, la moitié des pays ont entre 32.6 et 364.5 téléphones pour mille habitants.

Nous allons pour l'instant garder l'observation "hors norme" mais nous y reviendrons possiblement dessus dans d'autres parties du rapport.

Avant toutes recherches plus précises sur notre variable T , nous allons nous convaincre qu'elle est continue. Pour ce faire, traçons sa fonction de répartition empirique.



On voit que la fonction de répartition empirique de T réalise beaucoup de sauts. Elle en réalise exactement 200 qui est “grand” devant 209 donc notre variable d’intérêt est continue. Traçons son histogramme.

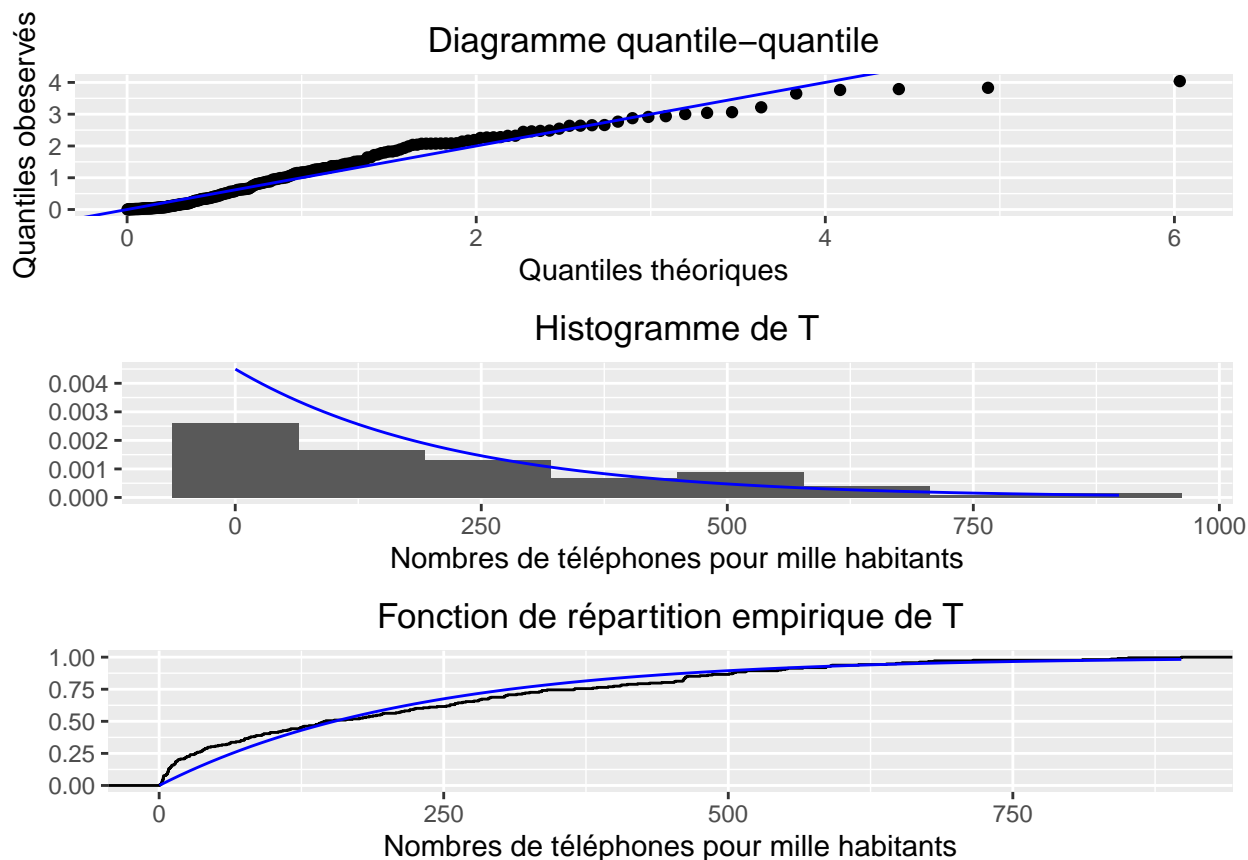


Les histogrammes sont uni-modaux avec un pic dans en 0, puis décroissant. On en déduit que dans notre jeu de données, beaucoup de pays ont moins de 100 téléphones pour mille habitants et à l'inverse, peu de pays on plus de 700 téléphones pour mille habitants.

Au vue de ces histogrammes, nous pouvons penser que T est un échantillon de loi exponentielle, c'est-à-dire X suit une loi exponentielle de paramètre λ . L'espérance d'une loi exponentielle est $\frac{1}{\lambda}$. La méthode des moments nous propose la moyenne empirique de T comme estimateur sans biais, consistant et asymptotiquement normal de l'espérance.

Dans le sommaire plus haut, la moyenne empirique de T est $\mu = 222.2533654$. En prenant $\hat{\lambda} = 0.0044994$, nous voulons vérifier que notre hypothèse est acceptable c'est à dire que la variable Téléphones suit une loi exponentielle de paramètre $\hat{\lambda} = 0.0044994$. Nous allons alors tracer:

1. L'histogramme de T et la densité $\hat{\lambda}e^{\hat{\lambda}x}$
2. La fonction de répartition empirique de T et la fonction de répartition de $X \sim \mathcal{E}(\hat{\lambda})$.
3. Le QQ-plot entre les quantiles théoriques de X et les quantiles empiriques de T .



Tout d'abord, les points de notre QQ-plot s'alignent bien sur la première bissectrice. C'est encourageant. De même, notre histogramme approche la courbe de densité de notre loi théorique et notre fonction de répartition fait de même. L'adéquation n'est pas parfaite mais elle est suffisamment raisonnable pour retenir l'hypothèse que T suit une loi exponentielle de paramètre $\hat{\lambda}$.

Nous allons finaliser notre analyse uni-variée en réalisant un test de Kolmogorov-Smirnov sur notre échantillon, puisque la variables Téléphones et loi exponentielle sont toutes deux des lois continues.

```
Warning in ks.test(df.countries$Téléphones, rexp/sample): p-value will be
approximate in the presence of ties
```

Notre p-valeur est 0 qui est inférieure à 0.01 donc on rejette H_0 au profit de l'hypothèse alternative H_1 sans hésitation. On peut conclure que notre variable T ne suit pas une loi exponentielle de paramètre $\hat{\lambda}$. T est la réalisation d'une autre loi, dont on ne connaît pour l'instant pas la nature exacte.

Nous avons donc terminé cette partie d'analyse univariée. Si le besoin s'en fait sentir dans des parties à venir, nous réaliserons d'autre test d'adéquation à des lois connues pour obtenir la loi de T .

C - Analyse bi-variée Nous allons dans cette partie nous intéresser aux potentielles corrélation entre notre variables d'intérêt T et les autres variables de notre jeu de données.

Commençons par calculer la matrice de corrélation et présentons un diagramme de dispersion de toutes les paires de variables, pour chaque variable quantitative de notre jeu de données.

	Téléphones	Densité.population	PIB
Afghanistan	FALSE	FALSE	FALSE
Albania	FALSE	FALSE	FALSE
Algeria	FALSE	FALSE	FALSE
Angola	FALSE	FALSE	FALSE

Anguilla	FALSE	FALSE FALSE
Antigua & Barbuda	FALSE	FALSE FALSE
Argentina	FALSE	FALSE FALSE
Armenia	FALSE	FALSE FALSE
Aruba	FALSE	FALSE FALSE
Australia	FALSE	FALSE FALSE
Austria	FALSE	FALSE FALSE
Azerbaijan	FALSE	FALSE FALSE
Bahamas, The	FALSE	FALSE FALSE
Bahrain	FALSE	FALSE FALSE
Bangladesh	FALSE	FALSE FALSE
Barbados	FALSE	FALSE FALSE
Belarus	FALSE	FALSE FALSE
Belgium	FALSE	FALSE FALSE
Belize	FALSE	FALSE FALSE
Benin	FALSE	FALSE FALSE
Bermuda	FALSE	FALSE FALSE
Bhutan	FALSE	FALSE FALSE
Bolivia	FALSE	FALSE FALSE
Bosnia & Herzegovina	FALSE	FALSE FALSE
Botswana	FALSE	FALSE FALSE
Brazil	FALSE	FALSE FALSE
British Virgin Is.	FALSE	FALSE FALSE
Brunei	FALSE	FALSE FALSE
Bulgaria	FALSE	FALSE FALSE
Burkina Faso	FALSE	FALSE FALSE
Burma	FALSE	FALSE FALSE
Burundi	FALSE	FALSE FALSE
Cambodia	FALSE	FALSE FALSE
Cameroon	FALSE	FALSE FALSE
Canada	FALSE	FALSE FALSE
Cape Verde	FALSE	FALSE FALSE
Cayman Islands	FALSE	FALSE FALSE
Central African Rep.	FALSE	FALSE FALSE
Chad	FALSE	FALSE FALSE
Chile	FALSE	FALSE FALSE
China	FALSE	FALSE FALSE
Colombia	FALSE	FALSE FALSE
Comoros	FALSE	FALSE FALSE
Congo, Dem. Rep.	TRUE	FALSE FALSE
Congo, Repub. of the	FALSE	FALSE FALSE
Cook Islands	FALSE	FALSE FALSE
Costa Rica	FALSE	FALSE FALSE
Cote d'Ivoire	FALSE	FALSE FALSE
Croatia	FALSE	FALSE FALSE
Cuba	FALSE	FALSE FALSE
Czech Republic	FALSE	FALSE FALSE
Denmark	FALSE	FALSE FALSE
Djibouti	FALSE	FALSE FALSE
Dominica	FALSE	FALSE FALSE
Dominican Republic	FALSE	FALSE FALSE
Ecuador	FALSE	FALSE FALSE
Egypt	FALSE	FALSE FALSE
El Salvador	FALSE	FALSE FALSE

Equatorial Guinea	FALSE	FALSE FALSE
Eritrea	FALSE	FALSE FALSE
Estonia	FALSE	FALSE FALSE
Ethiopia	FALSE	FALSE FALSE
Faroe Islands	FALSE	FALSE FALSE
Fiji	FALSE	FALSE FALSE
Finland	FALSE	FALSE FALSE
France	FALSE	FALSE FALSE
French Guiana	FALSE	FALSE FALSE
French Polynesia	FALSE	FALSE FALSE
Gabon	FALSE	FALSE FALSE
Gambia, The	FALSE	FALSE FALSE
Gaza Strip	FALSE	FALSE FALSE
Georgia	FALSE	FALSE FALSE
Germany	FALSE	FALSE FALSE
Ghana	FALSE	FALSE FALSE
Greece	FALSE	FALSE FALSE
Grenada	FALSE	FALSE FALSE
Guadeloupe	FALSE	FALSE FALSE
Guatemala	FALSE	FALSE FALSE
Guernsey	FALSE	FALSE FALSE
Guinea	FALSE	FALSE FALSE
Guinea-Bissau	FALSE	FALSE FALSE
Guyana	FALSE	FALSE FALSE
Haiti	FALSE	FALSE FALSE
Honduras	FALSE	FALSE FALSE
Hong Kong	FALSE	FALSE FALSE
Hungary	FALSE	FALSE FALSE
Iceland	FALSE	FALSE FALSE
India	FALSE	FALSE FALSE
Indonesia	FALSE	FALSE FALSE
Iran	FALSE	FALSE FALSE
Iraq	FALSE	FALSE FALSE
Ireland	FALSE	FALSE FALSE
Isle of Man	FALSE	FALSE FALSE
Israel	FALSE	FALSE FALSE
Italy	FALSE	FALSE FALSE
Jamaica	FALSE	FALSE FALSE
Japan	FALSE	FALSE FALSE
Jersey	FALSE	FALSE FALSE
Jordan	FALSE	FALSE FALSE
Kazakhstan	FALSE	FALSE FALSE
Kenya	FALSE	FALSE FALSE
Kiribati	FALSE	FALSE FALSE
Korea, North	FALSE	FALSE FALSE
Korea, South	FALSE	FALSE FALSE
Kuwait	FALSE	FALSE FALSE
Kyrgyzstan	FALSE	FALSE FALSE
Laos	FALSE	FALSE FALSE
Latvia	FALSE	FALSE FALSE
Lebanon	FALSE	FALSE FALSE
Lesotho	FALSE	FALSE FALSE
Liberia	FALSE	FALSE FALSE
Libya	FALSE	FALSE FALSE

Liechtenstein	FALSE	FALSE FALSE
Lithuania	FALSE	FALSE FALSE
Luxembourg	FALSE	FALSE FALSE
Macau	FALSE	FALSE FALSE
Macedonia	FALSE	FALSE FALSE
Madagascar	FALSE	FALSE FALSE
Malawi	FALSE	FALSE FALSE
Malaysia	FALSE	FALSE FALSE
Maldives	FALSE	FALSE FALSE
Mali	FALSE	FALSE FALSE
Malta	FALSE	FALSE FALSE
Marshall Islands	FALSE	FALSE FALSE
Martinique	FALSE	FALSE FALSE
Mauritania	FALSE	FALSE FALSE
Mauritius	FALSE	FALSE FALSE
Mexico	FALSE	FALSE FALSE
Micronesia, Fed. St.	FALSE	FALSE FALSE
Moldova	FALSE	FALSE FALSE
Mongolia	FALSE	FALSE FALSE
Morocco	FALSE	FALSE FALSE
Mozambique	FALSE	FALSE FALSE
Namibia	FALSE	FALSE FALSE
Nepal	FALSE	FALSE FALSE
Netherlands	FALSE	FALSE FALSE
Netherlands Antilles	FALSE	FALSE FALSE
New Caledonia	FALSE	FALSE FALSE
New Zealand	FALSE	FALSE FALSE
Nicaragua	FALSE	FALSE FALSE
Niger	FALSE	FALSE FALSE
Nigeria	FALSE	FALSE FALSE
Norway	FALSE	FALSE FALSE
Oman	FALSE	FALSE FALSE
Pakistan	FALSE	FALSE FALSE
Palau	FALSE	FALSE FALSE
Panama	FALSE	FALSE FALSE
Papua New Guinea	FALSE	FALSE FALSE
Paraguay	FALSE	FALSE FALSE
Peru	FALSE	FALSE FALSE
Philippines	FALSE	FALSE FALSE
Poland	FALSE	FALSE FALSE
Portugal	FALSE	FALSE FALSE
Puerto Rico	FALSE	FALSE FALSE
Qatar	FALSE	FALSE FALSE
Reunion	FALSE	FALSE FALSE
Romania	FALSE	FALSE FALSE
Russia	FALSE	FALSE FALSE
Rwanda	FALSE	FALSE FALSE
Saint Kitts & Nevis	FALSE	FALSE FALSE
Saint Lucia	FALSE	FALSE FALSE
Saint Vincent and the Grenadines	FALSE	FALSE FALSE
Samoa	FALSE	FALSE FALSE
Sao Tome & Principe	FALSE	FALSE FALSE
Saudi Arabia	FALSE	FALSE FALSE
Senegal	FALSE	FALSE FALSE

Seychelles	FALSE	FALSE	FALSE
Sierra Leone	FALSE	FALSE	FALSE
Singapore	FALSE	FALSE	FALSE
Slovakia	FALSE	FALSE	FALSE
Slovenia	FALSE	FALSE	FALSE
Solomon Islands	FALSE	FALSE	FALSE
Somalia	FALSE	FALSE	FALSE
South Africa	FALSE	FALSE	FALSE
Spain	FALSE	FALSE	FALSE
Sri Lanka	FALSE	FALSE	FALSE
Sudan	FALSE	FALSE	FALSE
Suriname	FALSE	FALSE	FALSE
Swaziland	FALSE	FALSE	FALSE
Sweden	FALSE	FALSE	FALSE
Switzerland	FALSE	FALSE	FALSE
Syria	FALSE	FALSE	FALSE
Taiwan	FALSE	FALSE	FALSE
Tajikistan	FALSE	FALSE	FALSE
Tanzania	FALSE	FALSE	FALSE
Thailand	FALSE	FALSE	FALSE
Togo	FALSE	FALSE	FALSE
Tonga	FALSE	FALSE	FALSE
Trinidad & Tobago	FALSE	FALSE	FALSE
Tunisia	FALSE	FALSE	FALSE
Turkey	FALSE	FALSE	FALSE
Turkmenistan	FALSE	FALSE	FALSE
Tuvalu	FALSE	FALSE	FALSE
Uganda	FALSE	FALSE	FALSE
Ukraine	FALSE	FALSE	FALSE
United Arab Emirates	FALSE	FALSE	FALSE
United Kingdom	FALSE	FALSE	FALSE
United States	FALSE	FALSE	FALSE
Uruguay	FALSE	FALSE	FALSE
Uzbekistan	FALSE	FALSE	FALSE
Vanuatu	FALSE	FALSE	FALSE
Venezuela	FALSE	FALSE	FALSE
Vietnam	FALSE	FALSE	FALSE
Virgin Islands	FALSE	FALSE	FALSE
West Bank	FALSE	FALSE	FALSE
Yemen	FALSE	FALSE	FALSE
Zambia	FALSE	FALSE	FALSE
Zimbabwe	FALSE	FALSE	FALSE

Agriculture Industrie Services

Afghanistan	FALSE	TRUE	TRUE
Albania	FALSE	TRUE	FALSE
Algeria	TRUE	FALSE	TRUE
Angola	TRUE	FALSE	TRUE
Anguilla	TRUE	TRUE	FALSE
Antigua & Barbuda	TRUE	TRUE	FALSE
Argentina	TRUE	FALSE	FALSE
Armenia	FALSE	FALSE	TRUE
Aruba	TRUE	FALSE	FALSE
Australia	TRUE	FALSE	FALSE
Austria	TRUE	FALSE	FALSE

Azerbaijan	FALSE	FALSE	TRUE
Bahamas, The	TRUE	TRUE	FALSE
Bahrain	TRUE	FALSE	FALSE
Bangladesh	FALSE	TRUE	FALSE
Barbados	TRUE	TRUE	FALSE
Belarus	TRUE	FALSE	FALSE
Belgium	TRUE	TRUE	FALSE
Belize	FALSE	TRUE	FALSE
Benin	FALSE	TRUE	FALSE
Bermuda	TRUE	TRUE	FALSE
Bhutan	FALSE	FALSE	TRUE
Bolivia	TRUE	FALSE	FALSE
Bosnia & Herzegovina	FALSE	FALSE	FALSE
Botswana	TRUE	FALSE	FALSE
Brazil	TRUE	FALSE	FALSE
British Virgin Is.	TRUE	TRUE	FALSE
Brunei	TRUE	FALSE	TRUE
Bulgaria	TRUE	FALSE	FALSE
Burkina Faso	FALSE	TRUE	TRUE
Burma	FALSE	TRUE	TRUE
Burundi	FALSE	TRUE	TRUE
Cambodia	FALSE	FALSE	TRUE
Cameroon	FALSE	TRUE	TRUE
Canada	TRUE	FALSE	FALSE
Cape Verde	TRUE	TRUE	FALSE
Cayman Islands	TRUE	TRUE	FALSE
Central African Rep.	FALSE	TRUE	TRUE
Chad	FALSE	FALSE	TRUE
Chile	TRUE	FALSE	TRUE
China	TRUE	FALSE	TRUE
Colombia	TRUE	FALSE	FALSE
Comoros	FALSE	TRUE	FALSE
Congo, Dem. Rep.	FALSE	TRUE	TRUE
Congo, Repub. of the	TRUE	FALSE	TRUE
Cook Islands	FALSE	TRUE	FALSE
Costa Rica	TRUE	FALSE	FALSE
Cote d'Ivoire	FALSE	TRUE	FALSE
Croatia	TRUE	FALSE	FALSE
Cuba	TRUE	FALSE	FALSE
Czech Republic	TRUE	FALSE	FALSE
Denmark	TRUE	TRUE	FALSE
Djibouti	FALSE	TRUE	FALSE
Dominica	FALSE	FALSE	FALSE
Dominican Republic	TRUE	FALSE	FALSE
Ecuador	TRUE	FALSE	FALSE
Egypt	FALSE	FALSE	FALSE
El Salvador	TRUE	FALSE	FALSE
Equatorial Guinea	TRUE	FALSE	TRUE
Eritrea	TRUE	FALSE	FALSE
Estonia	TRUE	FALSE	FALSE
Ethiopia	FALSE	TRUE	TRUE
Faroe Islands	FALSE	TRUE	FALSE
Fiji	TRUE	TRUE	FALSE
Finland	TRUE	FALSE	FALSE

France	TRUE	TRUE	FALSE
French Guiana	TRUE	TRUE	FALSE
French Polynesia	TRUE	TRUE	FALSE
Gabon	TRUE	FALSE	TRUE
Gambia, The	FALSE	TRUE	FALSE
Gaza Strip	TRUE	FALSE	FALSE
Georgia	FALSE	FALSE	FALSE
Germany	TRUE	FALSE	FALSE
Ghana	FALSE	TRUE	TRUE
Greece	TRUE	TRUE	FALSE
Grenada	TRUE	TRUE	FALSE
Guadeloupe	FALSE	TRUE	FALSE
Guatemala	FALSE	TRUE	FALSE
Guernsey	TRUE	TRUE	FALSE
Guinea	FALSE	FALSE	TRUE
Guinea-Bissau	FALSE	TRUE	TRUE
Guyana	FALSE	TRUE	TRUE
Haiti	FALSE	TRUE	FALSE
Honduras	FALSE	FALSE	FALSE
Hong Kong	TRUE	TRUE	FALSE
Hungary	TRUE	FALSE	FALSE
Iceland	TRUE	TRUE	FALSE
India	FALSE	FALSE	FALSE
Indonesia	FALSE	FALSE	TRUE
Iran	TRUE	FALSE	TRUE
Iraq	TRUE	FALSE	TRUE
Ireland	TRUE	FALSE	FALSE
Isle of Man	TRUE	TRUE	FALSE
Israel	TRUE	FALSE	FALSE
Italy	TRUE	FALSE	FALSE
Jamaica	TRUE	FALSE	FALSE
Japan	TRUE	FALSE	FALSE
Jersey	TRUE	TRUE	FALSE
Jordan	TRUE	FALSE	FALSE
Kazakhstan	TRUE	FALSE	FALSE
Kenya	FALSE	TRUE	FALSE
Kiribati	TRUE	TRUE	FALSE
Korea, North	FALSE	FALSE	TRUE
Korea, South	TRUE	FALSE	FALSE
Kuwait	TRUE	FALSE	FALSE
Kyrgyzstan	FALSE	TRUE	TRUE
Laos	FALSE	FALSE	TRUE
Latvia	TRUE	FALSE	FALSE
Lebanon	TRUE	TRUE	FALSE
Lesotho	FALSE	FALSE	TRUE
Liberia	FALSE	TRUE	TRUE
Libya	TRUE	FALSE	TRUE
Liechtenstein	TRUE	FALSE	FALSE
Lithuania	TRUE	FALSE	FALSE
Luxembourg	TRUE	TRUE	FALSE
Macau	TRUE	TRUE	FALSE
Macedonia	TRUE	FALSE	FALSE
Madagascar	FALSE	TRUE	FALSE
Malawi	FALSE	TRUE	FALSE

Malaysia	TRUE	FALSE	TRUE
Maldives	FALSE	TRUE	FALSE
Mali	FALSE	TRUE	TRUE
Malta	TRUE	TRUE	FALSE
Marshall Islands	FALSE	TRUE	FALSE
Martinique	TRUE	TRUE	FALSE
Mauritania	FALSE	FALSE	TRUE
Mauritius	TRUE	FALSE	FALSE
Mexico	TRUE	FALSE	FALSE
Micronesia, Fed. St.	FALSE	TRUE	FALSE
Moldova	FALSE	TRUE	FALSE
Mongolia	FALSE	TRUE	FALSE
Morocco	FALSE	FALSE	TRUE
Mozambique	FALSE	FALSE	TRUE
Namibia	TRUE	FALSE	FALSE
Nepal	FALSE	TRUE	TRUE
Netherlands	TRUE	TRUE	FALSE
Netherlands Antilles	TRUE	TRUE	FALSE
New Caledonia	FALSE	TRUE	FALSE
New Zealand	TRUE	FALSE	FALSE
Nicaragua	FALSE	FALSE	FALSE
Niger	FALSE	TRUE	TRUE
Nigeria	FALSE	FALSE	TRUE
Norway	TRUE	FALSE	FALSE
Oman	TRUE	FALSE	FALSE
Pakistan	FALSE	FALSE	FALSE
Palau	TRUE	TRUE	FALSE
Panama	TRUE	TRUE	FALSE
Papua New Guinea	FALSE	FALSE	TRUE
Paraguay	FALSE	TRUE	FALSE
Peru	TRUE	FALSE	FALSE
Philippines	FALSE	FALSE	FALSE
Poland	TRUE	FALSE	FALSE
Portugal	TRUE	FALSE	FALSE
Puerto Rico	TRUE	FALSE	FALSE
Qatar	TRUE	FALSE	TRUE
Reunion	TRUE	TRUE	FALSE
Romania	TRUE	FALSE	FALSE
Russia	TRUE	FALSE	FALSE
Rwanda	FALSE	TRUE	TRUE
Saint Kitts & Nevis	TRUE	FALSE	FALSE
Saint Lucia	TRUE	TRUE	FALSE
Saint Vincent and the Grenadines	TRUE	FALSE	FALSE
Samoa	TRUE	FALSE	TRUE
Sao Tome & Principe	FALSE	TRUE	FALSE
Saudi Arabia	TRUE	FALSE	TRUE
Senegal	FALSE	TRUE	FALSE
Seychelles	TRUE	FALSE	FALSE
Sierra Leone	FALSE	FALSE	TRUE
Singapore	TRUE	FALSE	FALSE
Slovakia	TRUE	FALSE	FALSE
Slovenia	TRUE	FALSE	FALSE
Solomon Islands	FALSE	TRUE	TRUE
Somalia	FALSE	TRUE	TRUE

South Africa	TRUE	FALSE	FALSE
Spain	TRUE	FALSE	FALSE
Sri Lanka	FALSE	FALSE	FALSE
Sudan	FALSE	TRUE	TRUE
Suriname	TRUE	TRUE	FALSE
Swaziland	TRUE	FALSE	TRUE
Sweden	TRUE	FALSE	FALSE
Switzerland	TRUE	FALSE	FALSE
Syria	FALSE	TRUE	FALSE
Taiwan	TRUE	FALSE	FALSE
Tajikistan	FALSE	FALSE	TRUE
Tanzania	FALSE	TRUE	TRUE
Thailand	TRUE	FALSE	TRUE
Togo	FALSE	TRUE	TRUE
Tonga	FALSE	FALSE	FALSE
Trinidad & Tobago	TRUE	FALSE	TRUE
Tunisia	TRUE	FALSE	FALSE
Turkey	TRUE	FALSE	FALSE
Turkmenistan	FALSE	FALSE	TRUE
Tuvalu	FALSE	FALSE	FALSE
Uganda	FALSE	TRUE	TRUE
Ukraine	FALSE	FALSE	TRUE
United Arab Emirates	TRUE	FALSE	TRUE
United Kingdom	TRUE	TRUE	FALSE
United States	TRUE	TRUE	FALSE
Uruguay	TRUE	FALSE	FALSE
Uzbekistan	FALSE	TRUE	TRUE
Vanuatu	FALSE	TRUE	FALSE
Venezuela	TRUE	FALSE	FALSE
Vietnam	FALSE	FALSE	TRUE
Virgin Islands	TRUE	TRUE	FALSE
West Bank	TRUE	FALSE	FALSE
Yemen	FALSE	FALSE	TRUE
Zambia	FALSE	FALSE	FALSE
Zimbabwe	FALSE	TRUE	FALSE

Taux.de.natalité

Afghanistan	FALSE
Albania	FALSE
Algeria	FALSE
Angola	FALSE
Anguilla	FALSE
Antigua & Barbuda	FALSE
Argentina	FALSE
Armenia	FALSE
Aruba	FALSE
Australia	FALSE
Austria	FALSE
Azerbaijan	FALSE
Bahamas, The	FALSE
Bahrain	FALSE
Bangladesh	FALSE
Barbados	FALSE
Belarus	FALSE
Belgium	FALSE

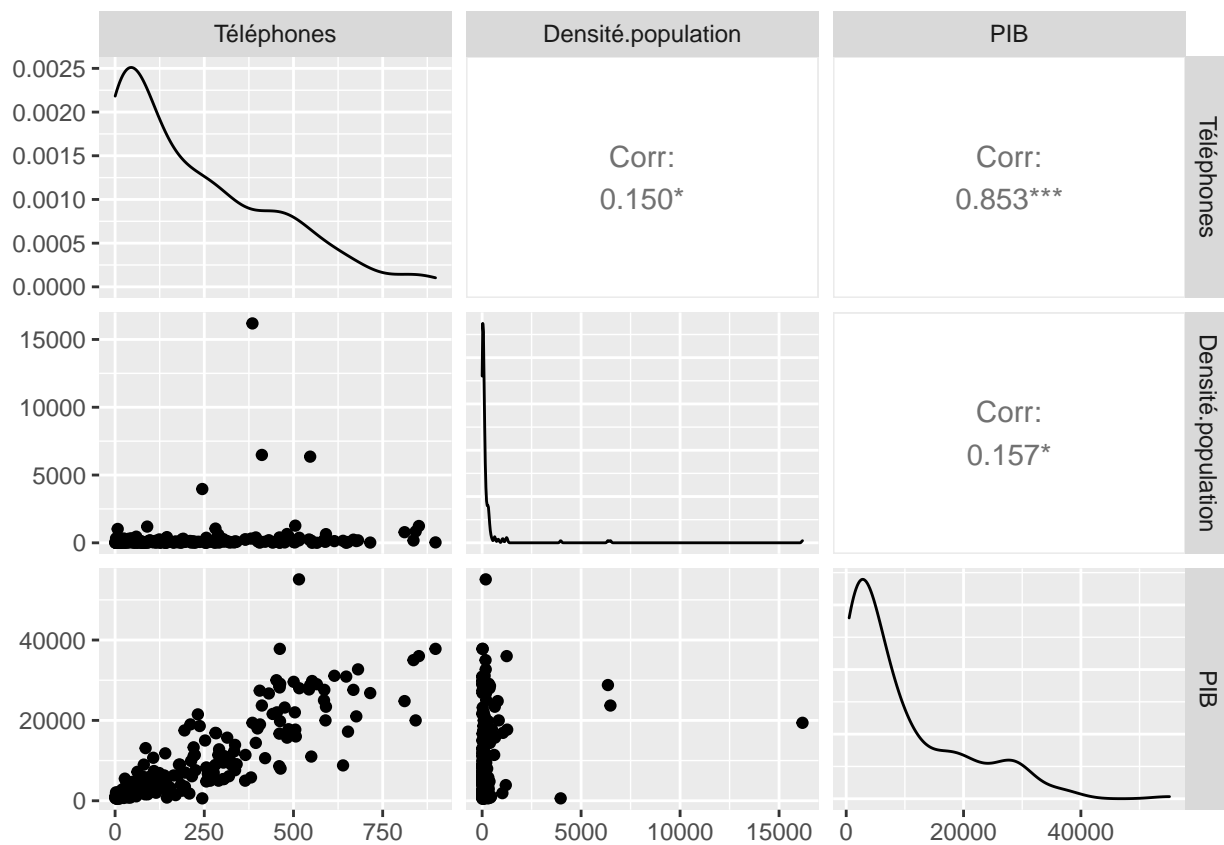
Belize	FALSE
Benin	FALSE
Bermuda	FALSE
Bhutan	FALSE
Bolivia	FALSE
Bosnia & Herzegovina	FALSE
Botswana	FALSE
Brazil	FALSE
British Virgin Is.	FALSE
Brunei	FALSE
Bulgaria	FALSE
Burkina Faso	FALSE
Burma	FALSE
Burundi	FALSE
Cambodia	FALSE
Cameroon	FALSE
Canada	FALSE
Cape Verde	FALSE
Cayman Islands	FALSE
Central African Rep.	FALSE
Chad	FALSE
Chile	FALSE
China	FALSE
Colombia	FALSE
Comoros	FALSE
Congo, Dem. Rep.	FALSE
Congo, Repub. of the	FALSE
Cook Islands	FALSE
Costa Rica	FALSE
Cote d'Ivoire	FALSE
Croatia	FALSE
Cuba	FALSE
Czech Republic	FALSE
Denmark	FALSE
Djibouti	FALSE
Dominica	FALSE
Dominican Republic	FALSE
Ecuador	FALSE
Egypt	FALSE
El Salvador	FALSE
Equatorial Guinea	FALSE
Eritrea	FALSE
Estonia	FALSE
Ethiopia	FALSE
Faroe Islands	FALSE
Fiji	FALSE
Finland	FALSE
France	FALSE
French Guiana	FALSE
French Polynesia	FALSE
Gabon	FALSE
Gambia, The	FALSE
Gaza Strip	FALSE
Georgia	FALSE

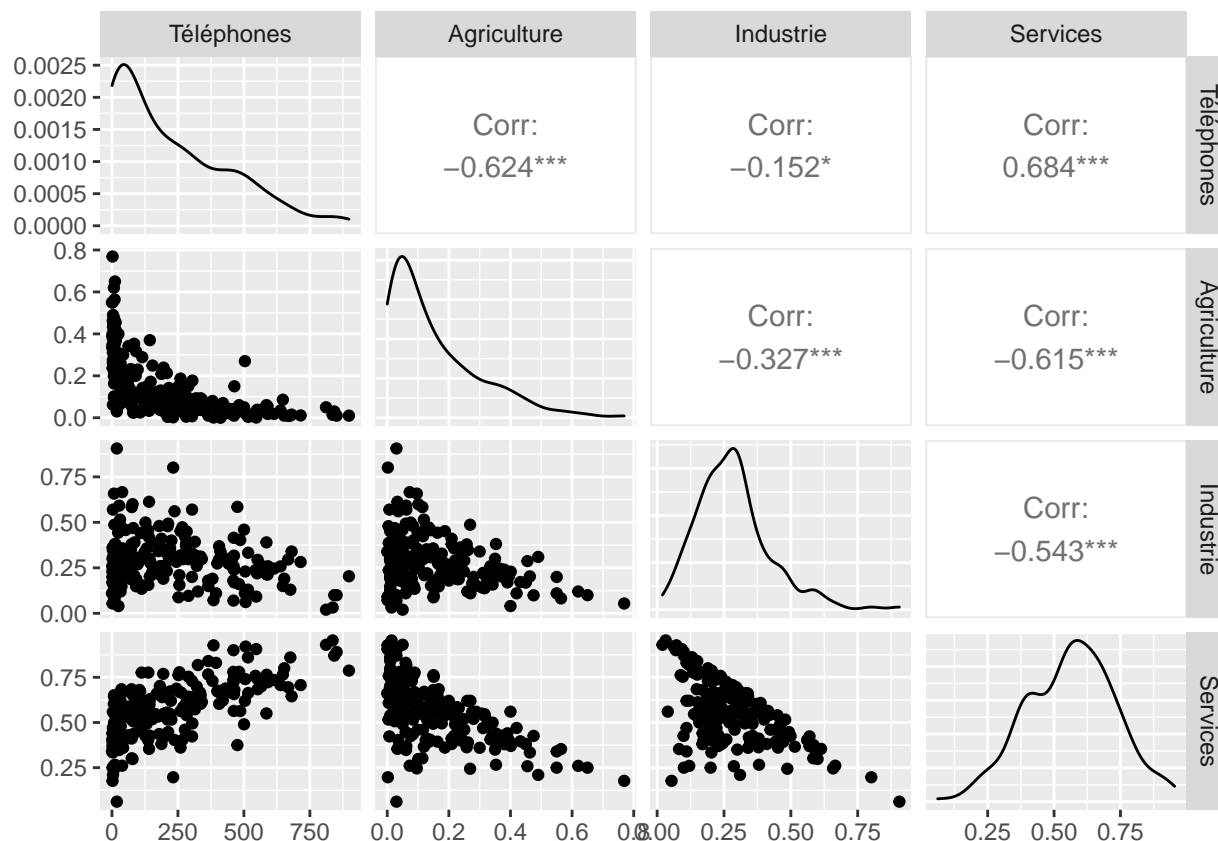
Germany	FALSE
Ghana	FALSE
Greece	FALSE
Grenada	FALSE
Guadeloupe	FALSE
Guatemala	FALSE
Guernsey	FALSE
Guinea	FALSE
Guinea-Bissau	FALSE
Guyana	FALSE
Haiti	FALSE
Honduras	FALSE
Hong Kong	FALSE
Hungary	FALSE
Iceland	FALSE
India	FALSE
Indonesia	FALSE
Iran	FALSE
Iraq	FALSE
Ireland	FALSE
Isle of Man	FALSE
Israel	FALSE
Italy	FALSE
Jamaica	FALSE
Japan	FALSE
Jersey	FALSE
Jordan	FALSE
Kazakhstan	FALSE
Kenya	FALSE
Kiribati	FALSE
Korea, North	FALSE
Korea, South	FALSE
Kuwait	FALSE
Kyrgyzstan	FALSE
Laos	FALSE
Latvia	FALSE
Lebanon	FALSE
Lesotho	FALSE
Liberia	FALSE
Libya	FALSE
Liechtenstein	FALSE
Lithuania	FALSE
Luxembourg	FALSE
Macau	FALSE
Macedonia	FALSE
Madagascar	FALSE
Malawi	FALSE
Malaysia	FALSE
Maldives	FALSE
Mali	FALSE
Malta	FALSE
Marshall Islands	FALSE
Martinique	FALSE
Mauritania	FALSE

Mauritius	FALSE
Mexico	FALSE
Micronesia, Fed. St.	FALSE
Moldova	FALSE
Mongolia	FALSE
Morocco	FALSE
Mozambique	FALSE
Namibia	FALSE
Nepal	FALSE
Netherlands	FALSE
Netherlands Antilles	FALSE
New Caledonia	FALSE
New Zealand	FALSE
Nicaragua	FALSE
Niger	FALSE
Nigeria	FALSE
Norway	FALSE
Oman	FALSE
Pakistan	FALSE
Palau	FALSE
Panama	FALSE
Papua New Guinea	FALSE
Paraguay	FALSE
Peru	FALSE
Philippines	FALSE
Poland	FALSE
Portugal	FALSE
Puerto Rico	FALSE
Qatar	FALSE
Reunion	FALSE
Romania	FALSE
Russia	FALSE
Rwanda	FALSE
Saint Kitts & Nevis	FALSE
Saint Lucia	FALSE
Saint Vincent and the Grenadines	FALSE
Samoa	FALSE
Sao Tome & Principe	FALSE
Saudi Arabia	FALSE
Senegal	FALSE
Seychelles	FALSE
Sierra Leone	FALSE
Singapore	FALSE
Slovakia	FALSE
Slovenia	FALSE
Solomon Islands	FALSE
Somalia	FALSE
South Africa	FALSE
Spain	FALSE
Sri Lanka	FALSE
Sudan	FALSE
Suriname	FALSE
Swaziland	FALSE
Sweden	FALSE

Switzerland	FALSE
Syria	FALSE
Taiwan	FALSE
Tajikistan	FALSE
Tanzania	FALSE
Thailand	FALSE
Togo	FALSE
Tonga	FALSE
Trinidad & Tobago	FALSE
Tunisia	FALSE
Turkey	FALSE
Turkmenistan	FALSE
Tuvalu	FALSE
Uganda	FALSE
Ukraine	FALSE
United Arab Emirates	FALSE
United Kingdom	FALSE
United States	FALSE
Uruguay	FALSE
Uzbekistan	FALSE
Vanuatu	FALSE
Venezuela	FALSE
Vietnam	FALSE
Virgin Islands	FALSE
West Bank	FALSE
Yemen	FALSE
Zambia	FALSE
Zimbabwe	FALSE

	Téléphones	Densité.population	PIB	Agriculture
Téléphones	1.0000000	0.1498945	0.85279091	-0.6241662
Densité.population	0.1498945	1.0000000	0.15678194	-0.1491919
PIB	0.8527909	0.1567819	1.00000000	-0.5944425
Agriculture	-0.6241662	-0.1491919	-0.59444251	1.0000000
Industrie	-0.1522105	-0.1409641	-0.02858728	-0.3269630
Services	0.6839644	0.2511465	0.55357338	-0.6151770
Taux.de.natalité	-0.7295612	-0.1399784	-0.64951839	0.6816288
	Industrie	Services	Taux.de.natalité	
Téléphones	-0.15221054	0.6839644	-0.72956121	
Densité.population	-0.14096408	0.2511465	-0.13997842	
PIB	-0.02858728	0.5535734	-0.64951839	
Agriculture	-0.32696302	-0.6151770	0.68162882	
Industrie	1.00000000	-0.5429759	-0.05900188	
Services	-0.54297594	1.0000000	-0.55789511	
Taux.de.natalité	-0.05900188	-0.5578951	1.00000000	





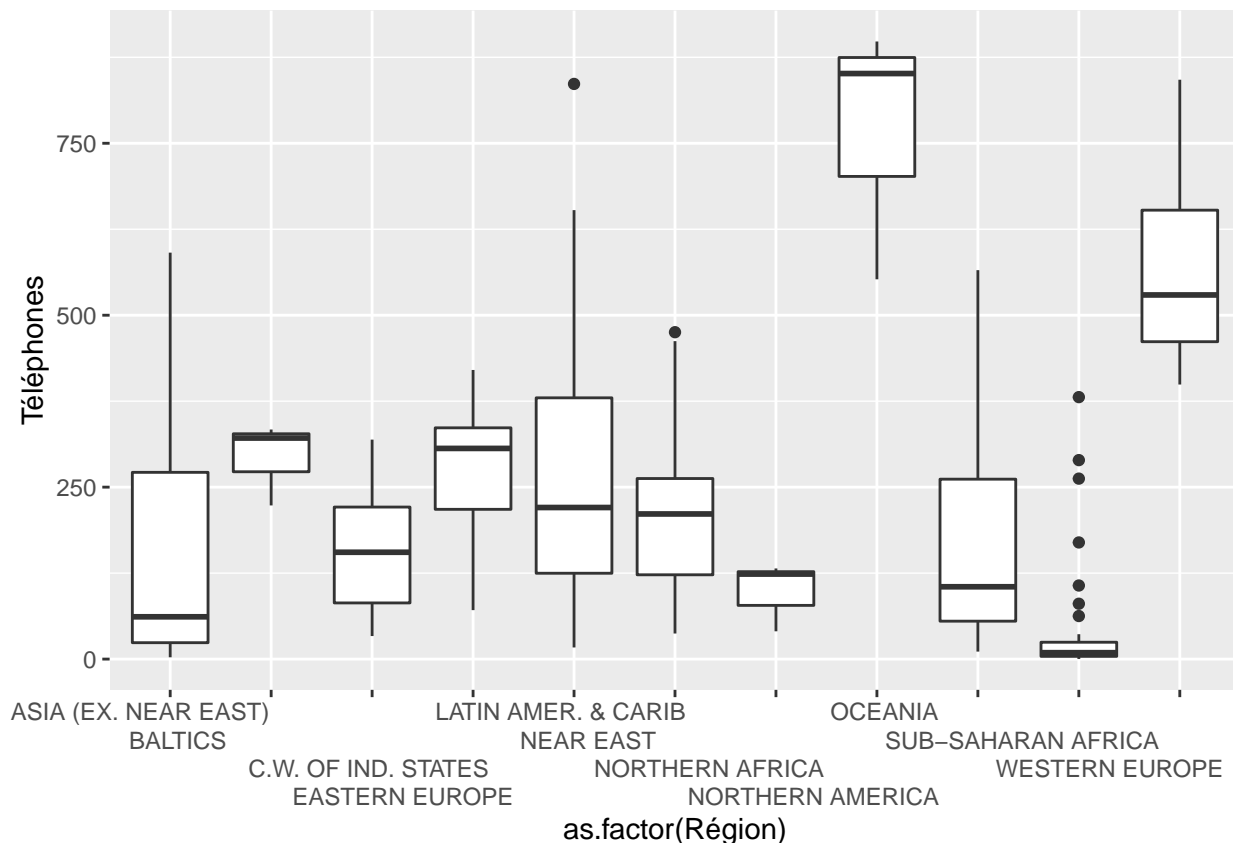
On a un facteur de corrélation d'environ 0.85 pour le couple de variables T et PIB; on en déduit qu'elles sont fortement corrélées linéairement. Il en est de même pour les variables Services et T avec un facteur de corrélation de 0.68. Finalement, la variable T est également corrélée négativement à la variable Agriculture avec un coefficient de corrélation de -0,62.

En analysant maintenant les nuages du points des paires de variables, on en déduit que les couples :

- (T , Densité de population): la tendance semble montrer une corrélation, mais c'est peut être l'effet de certaines observations très grandes, qui ont écrasé un nuage de points qui aurait semblé aléatoire. Le coefficient est en lui même faible.
- (T , PIB): les points sont relativement alignés sur une droite, ce qui était cohérent avec le coefficient de corrélation.
- (T , Agriculture): les points ont l'air concentrés dans une région du plan sous une hyperbole et dans le premier quart du plan. Le modèle idéal n'est peut être pas linéaire pour cette variable.
- (T , Industrie): le nuage de points semble un peu plus éparpillé, mais le coefficient de corrélation est nettement plus faible que pour les autres variables.
- (T , Services): ici la tendance est marquée sur le nuage, avec le plus grand des coefficients de corrélation.

Nous avons maintenant une idée plus claire des liens entre notre variables d'intérêt et les variables quantitatives de notre jeu de données.

Nous allons terminer cette partie en nous intéressant à la variables Région, qui est une variable qualitative. Traçons le boxplot de la variable T pour chaque modalité de la variable Région.



Les boxplots sont extrêmement différents, on peut penser que la région d'appartenance du pays joue un rôle important sur le nombre de téléphones pour mille habitants. On remarque qu'en Amérique du nord et en Europe de l'ouest, les médianes sont hautes alors qu'elles sont beaucoup plus basses en Afrique Sub-saharienne, en Océanie ou encore en Asie. Il y a parfois des observations hors-norme, notamment en Afrique sub-saharienne. Nous reviendrons sur ces observations plus loin dans le rapport.

II- Régression linéaire multiple

Dans cette partie, nous allons nous intéresser à une régression linéaire permettant d'exprimer le nombre de téléphone pour mille habitants par la densité de population (nombre habitants par miles carré), le PIB par habitant et enfin la part de la population active travaillant dans l'agriculture, l'industrie et les services.

On considère le modèle linéaire suivant:

$\text{Téléphones} = \beta_0 + \beta_1 \text{PIB} + \beta_2 \text{Densité} + \beta_3 \text{Agriculture} + \beta_4 \text{Industrie} + \beta_5 \text{Services} + \varepsilon$ Commençons par centrer et réduire les données pour qu'elles soient à la même échelle. Puis, réalisons une première régression linéaire.

Call:

```
lm(formula = Téléphones ~ Densité.population + PIB + Agriculture +
    Industrie + Services, data = df.countries)
```

Residuals:

Min	1Q	Median	3Q	Max
-485.38	-56.99	-5.64	46.16	360.79

Coefficients:

Estimate	Std. Error	t value	Pr(> t)

(Intercept)	-1.210e+03	1.049e+03	-1.153	0.250
Densité.population	-6.039e-03	5.457e-03	-1.107	0.270
PIB	1.430e-02	8.957e-04	15.965	<2e-16 ***
Agriculture	1.011e+03	1.049e+03	0.964	0.336
Industrie	1.095e+03	1.048e+03	1.044	0.298
Services	1.478e+03	1.053e+03	1.404	0.162

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

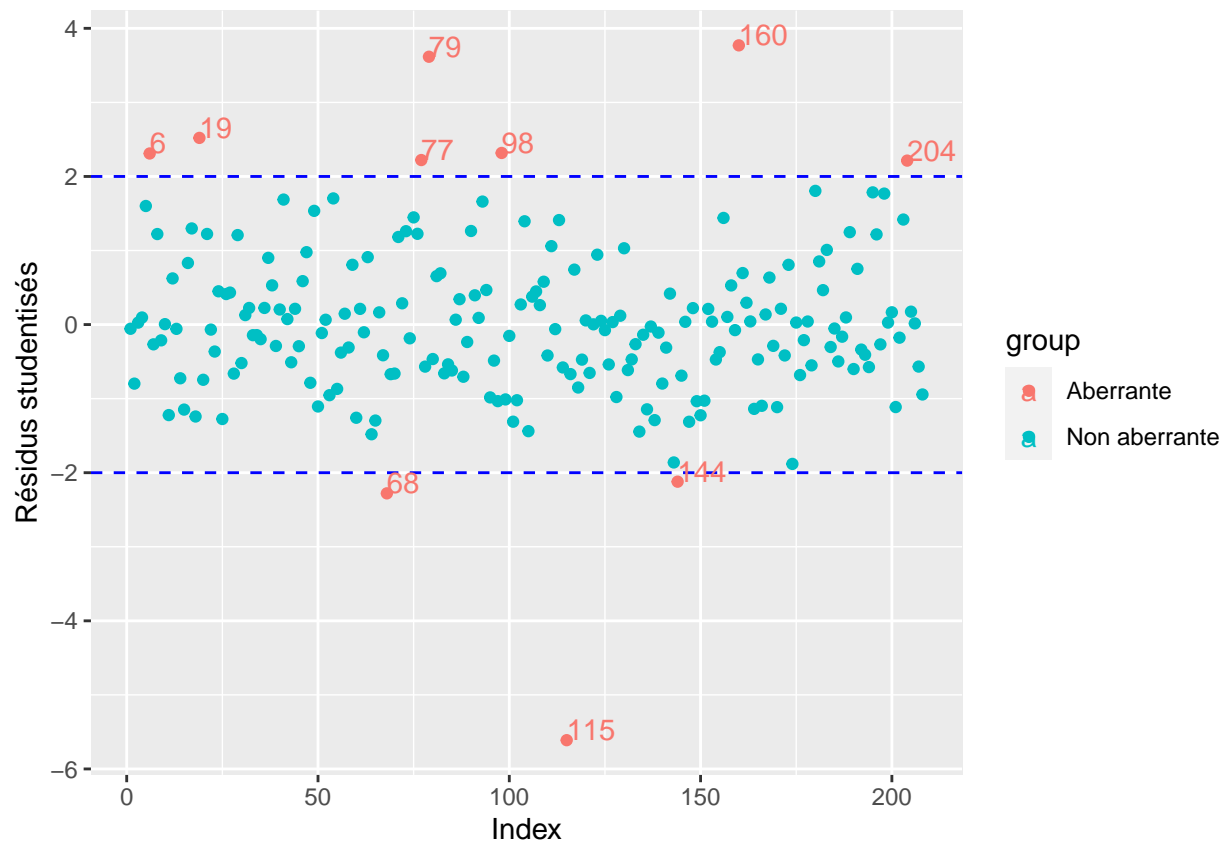
Residual standard error: 99.42 on 202 degrees of freedom

Multiple R-squared: 0.796, Adjusted R-squared: 0.7909

F-statistic: 157.6 on 5 and 202 DF, p-value: < 2.2e-16

Avant d'aller plus loin dans la régression linéaire et la sélection de variables nous allons nous intéresser à notre jeu de données, voir s'il comporte des valeurs aberrantes, des points leviers et analyser les distances de Cook.

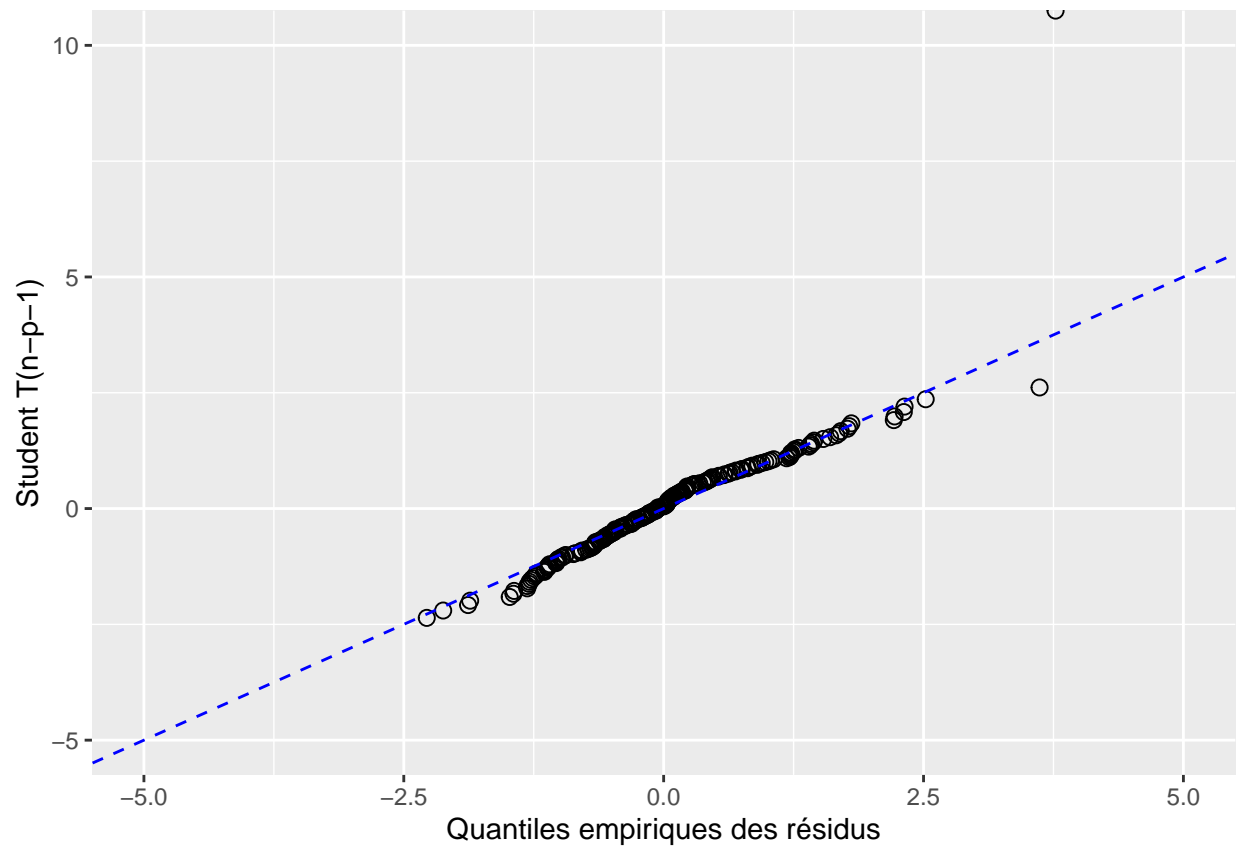
1) Valeurs aberrantes



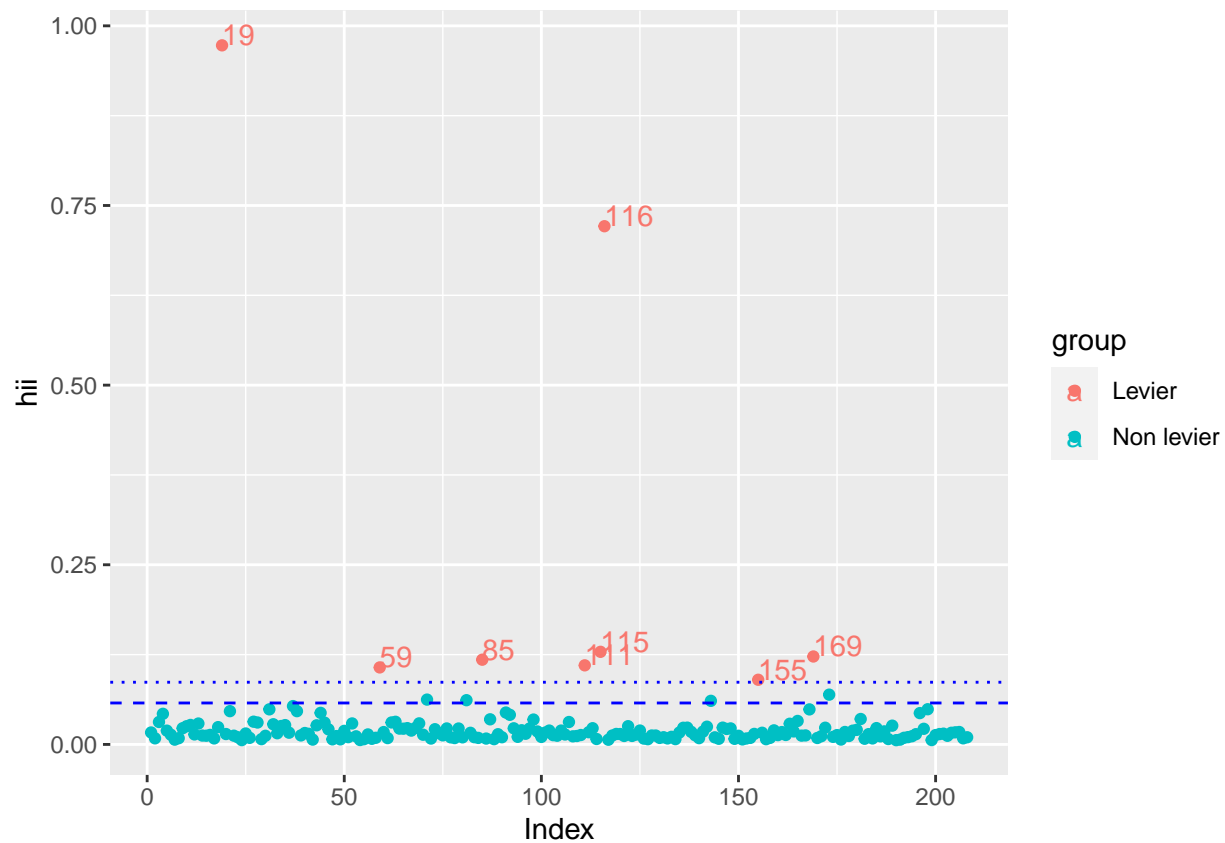
On a 10 valeurs aberrantes dans un échantillon de taille 209.

2) QQ-plot

Warning: Removed 1 rows containing missing values (geom_point).

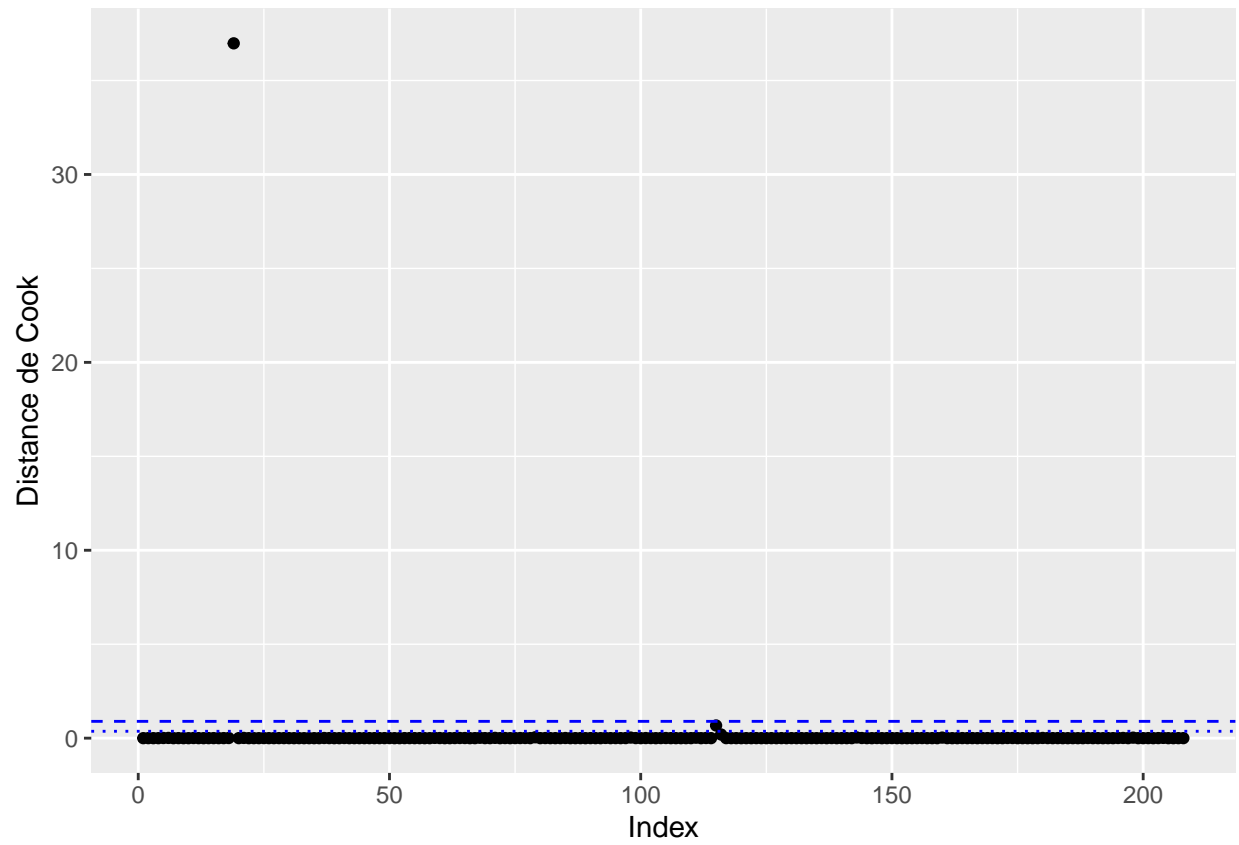


3) Points levier



Il y a 7 valeurs qui dépassent les deux seuils.

4) Distance de Cook



On observe qu'une valeur dépasse le seuil préoccupant du quantile $f_{n-p}^n(0.5)$.

[1] 0.9728572

[1] 36.97948

Belize

19

