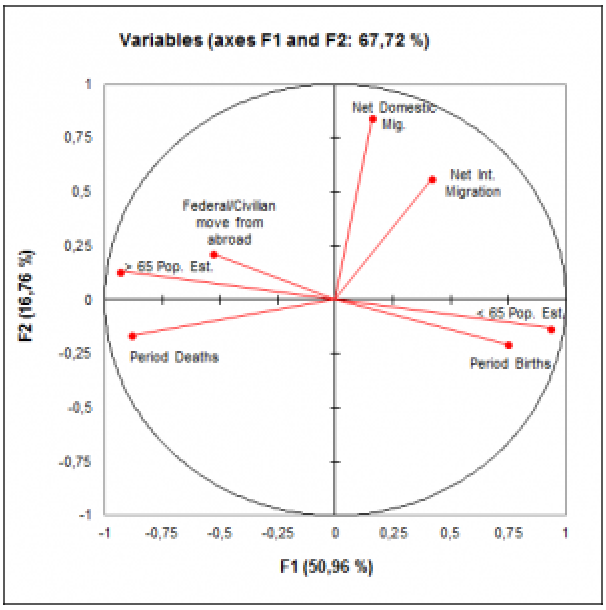
Wissam TAHIRI

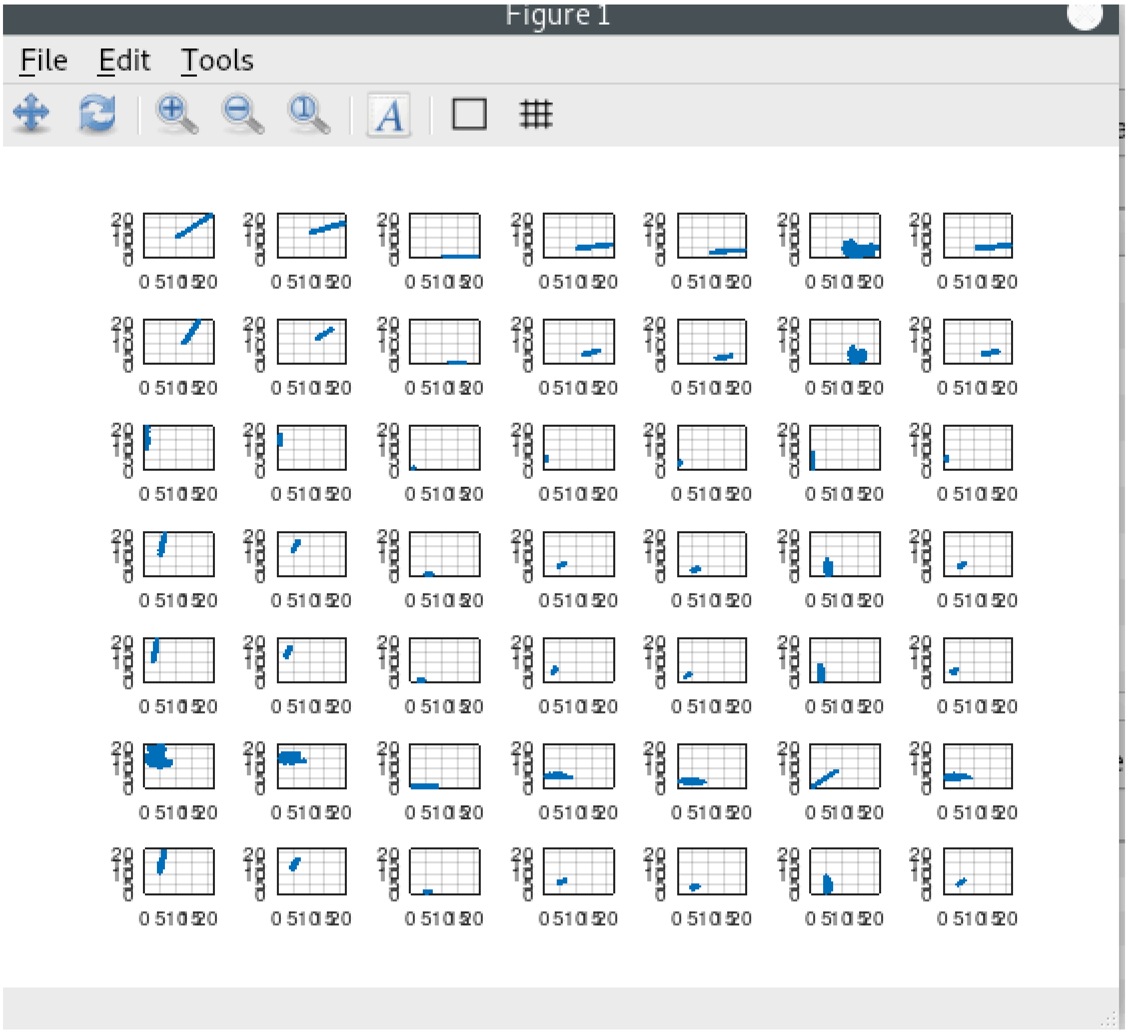
Rapport TP:
Analyser un jeu de données



Sébastien Guérif

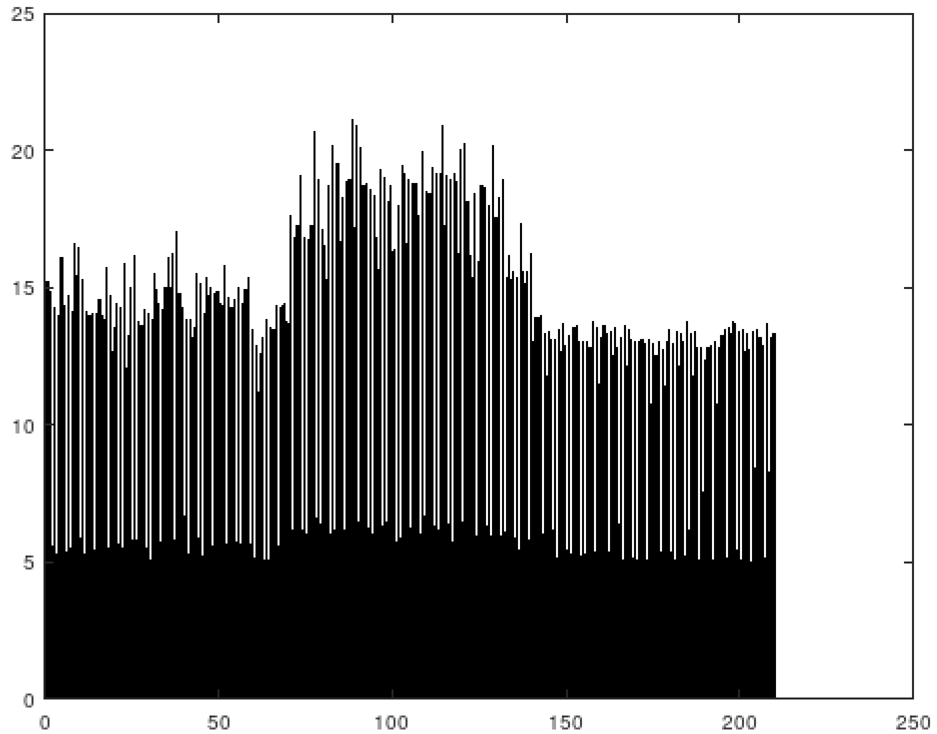
Analyse de données 23/05/2024

4) Voici ce que l'on obtient quand nous observons les variables les une par rapport aux autres par la matrice X:

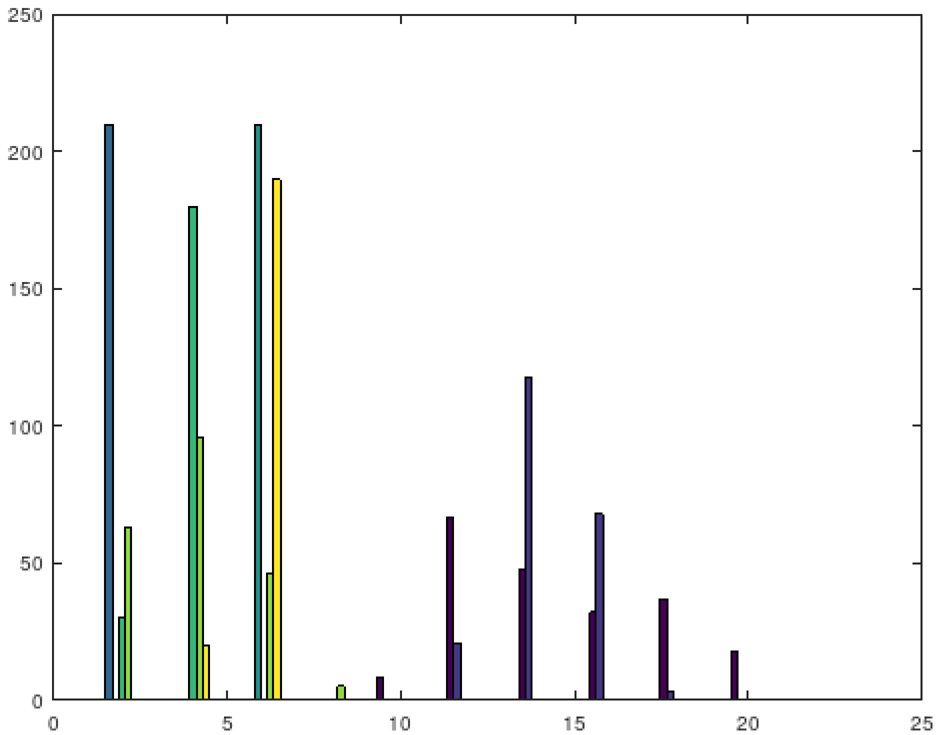


Et avec des fonctions caractéristiques comme:

- bar:

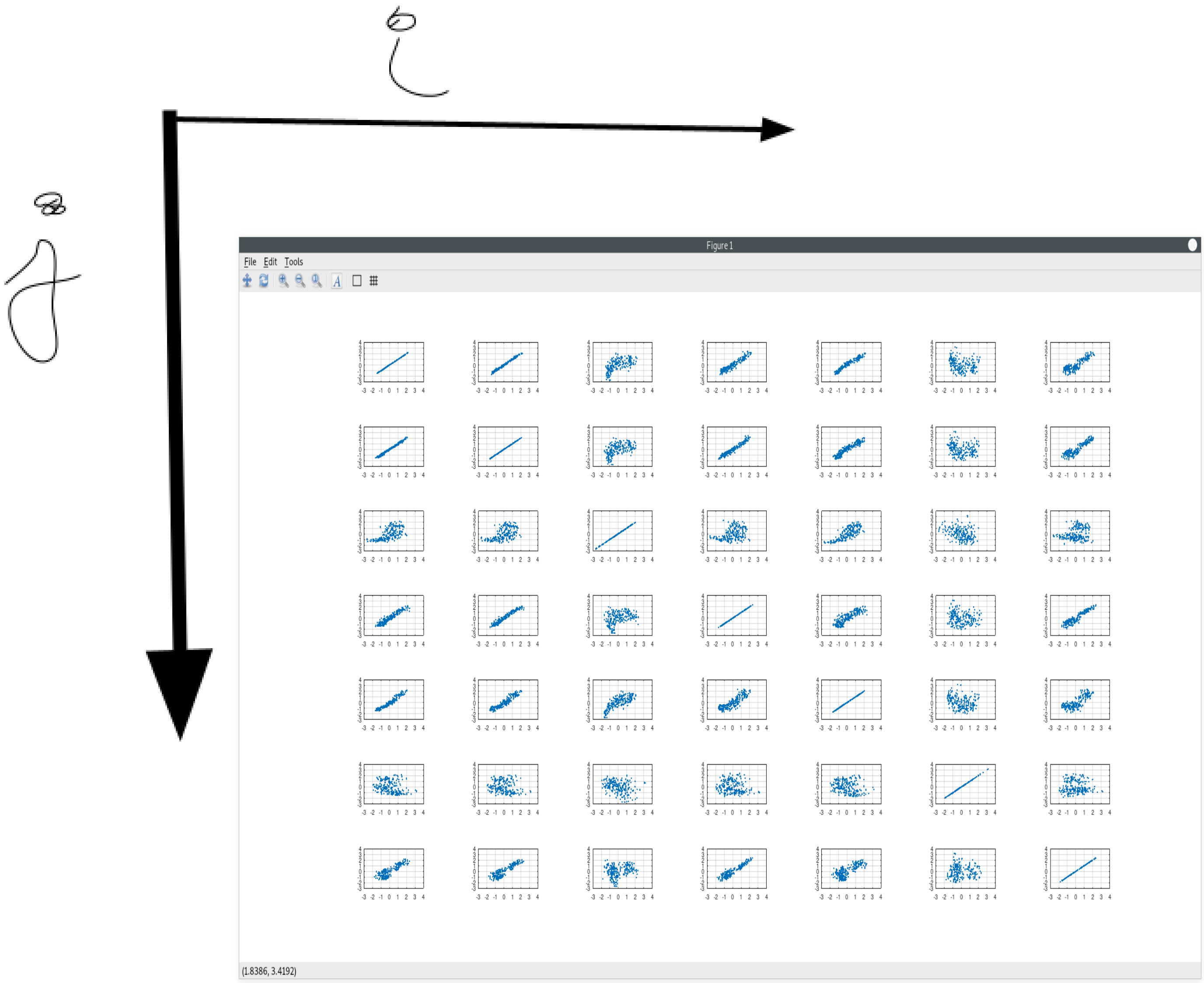


- hist:



entre autres.

6) Lorsque nous formons Xcr, nous obtenons une matrice centrée et réduite. Lorsque nous visualisons les variables les unes par rapport aux autres, nous obtenons ceci:

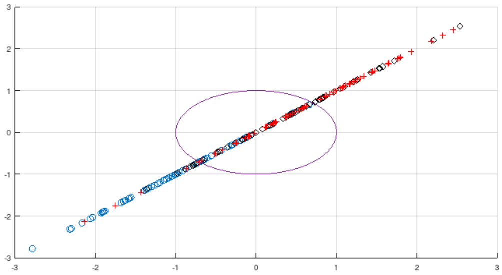
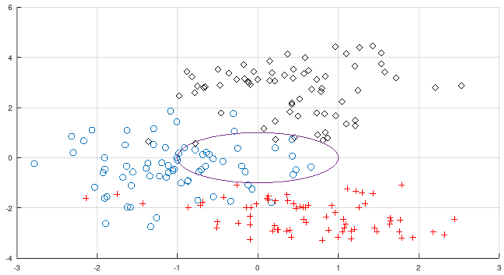


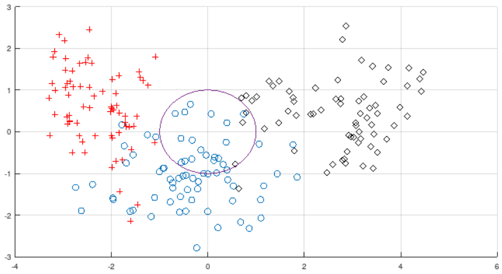
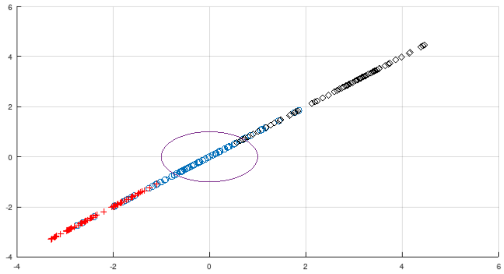
qui représente assez bien la corrélation entre les paramètres. En effet, corcoeff(X) donne

1.000000	0.994341	0.608288	0.949985	0.970771	-0.229572	0.863693
0.994341	1.000000	0.529244	0.972422	0.944829	-0.217340	0.890784
0.608288	0.529244	1.000000	0.367915	0.761635	-0.331471	0.226825
0.949985	0.972422	0.367915	1.000000	0.860415	-0.171562	0.932806
0.970771	0.944829	0.761635	0.860415	1.000000	-0.258037	0.749131
-0.229572	-0.217340	-0.331471	-0.171562	-0.258037	1.000000	-0.011079
0.863693	0.890784	0.226825	0.932806	0.749131	-0.011079	1.000000

Et plus la valeur est proche de 1, plus les paramètres sont corrélés. S'ils sont négatifs, ils sont inversement corrélés Comme nous pouvons le constater, si la valeur de X(i,j) est proche de 1 alors le graphe j=f(i) est approximativement linéaire comme nous pouvons l'observer pour les diagonales, (1;2), (1;4) etc…

7) Voici les graphes obtenus:

8) A l'aide d'un petit bout de code:

```
l = diag(lambda);
i_tot= sum(l);
prop= l / i_tot;
```

On obtient le vecteur prop qui contient les pourcentages d'apports d'Inertie à l'inertie totale par chaque composante. et on obtient alors ceci:

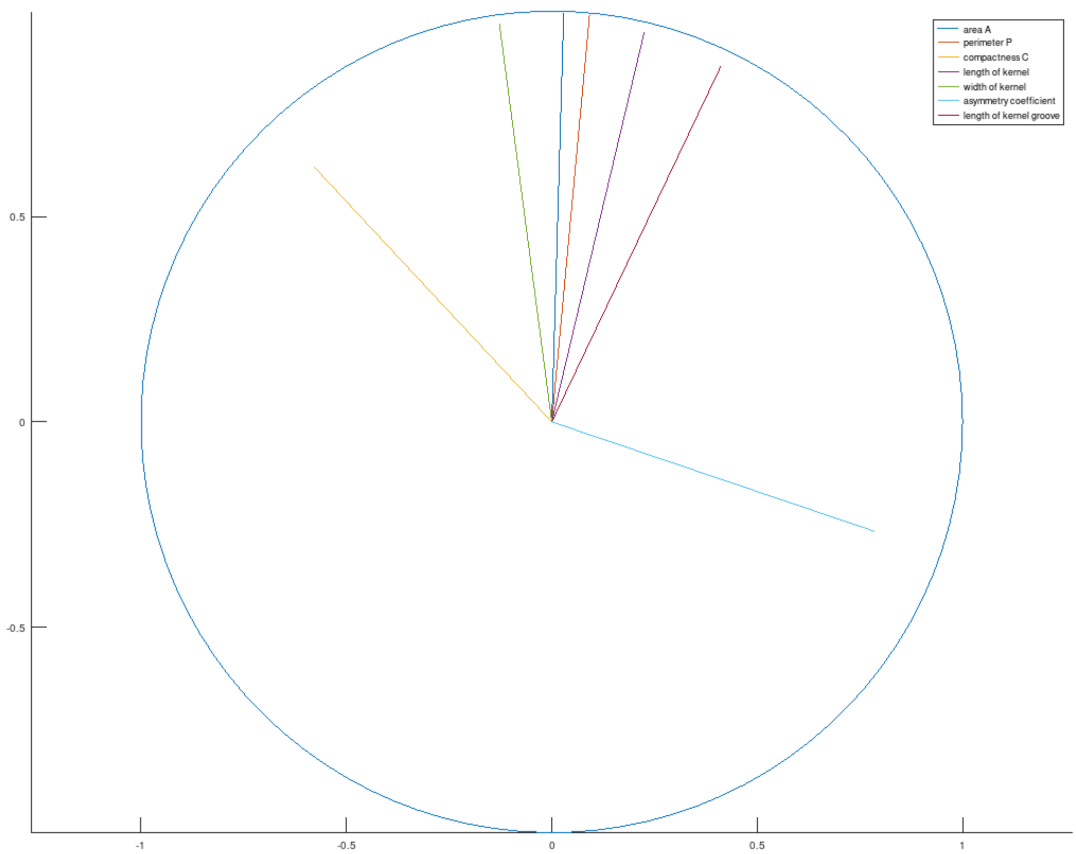
```
disp(prop)
1.1606e-04
7.6172e-04
2.6734e-03
9.7664e-03
9.6858e-02
1.7108e-01
7.1874e-01
```

Ainsi on peut voir que les 2 dernières variables apportent le plus d'inertie (89%) ainsi un plan formé par ces 2 composantes devrait afficher au mieux le graphique et afficher clairement les points sans trop les écraser. C'est également le plan des composantes dont les valeurs propres sont les plus grandes.

```
disp(l)
8.1240e-04
5.3320e-03
1.8714e-02
6.8364e-02
6.7800e-01
1.1976e+00
5.0312e+00
```

Ainsi la qualité de représentation des données projetées sur le premier plan factoriel est très bonne est ce plan est pertinent.

9)



Ainsi on peut voir que le coefficient d'asymétrie est assez corrélé à l'axe x tandis que les données de “taille” tels que l'épaisseur, la longueur, etc est assez corrélé à l'axe y. On peut également voir que la “compactness” est inversement corrélé à l'axe x.