

Travail personnel

Module Visualisation d'Information

Visualisation des informations pour d'une analyse médicale

Réalisé par

KHEDIM Ibtissem
BOUKEFTANE Bilal

Option: MST1

Année universitaire 2021/2022

Table des matières

1	Introduction	4
2	Processus de génération d'InfoVis	4
2.1	Analyse de données	5
2.2	Filtering	5
2.3	Mapping	7
2.4	Rendering	8
3	Conclusion	11

Table des figures

Figure 1 Nombre de fumeur et non-fumeur selon la l'activité professionnelle	6
Figure 2Pratique du sport par sexe	7
Figure 3 Répartition de la population par catégorie professionnelle	7
Figure 4 Répartition de la population par sexe	7
Figure 5 Répartition de la population par tranche d'âge	7
Figure 6 tree diagram représentant les statistiques de fumeur / Non-fumeur de selon le sexe et l'âge	8
Figure 7 diagramme de bars représentant Statistiques fumeur/Non-fumeur par activité professionnelle	9
Figure 8 diagramme de bars représentant Statistiques fumeur/Non-fumeur par sexe	9
Figure 9 diagramme circulaire représente la répartition de la population par tranche d'âge	10
Figure 10 diagramme circulaire représentant la répartition de la population par sexe	10
Figure 11 diagramme circulaire représentant la répartition de la population par activité professionnelle	11

1 Introduction

La visualisation d'information est l'utilisation d'une représentation visuelle interactive et assistée par ordinateur de données abstraites pour amplifier la cognition [CARD, MACKINLAY et SHNEIDERMAN, 1999].

Le domaine a émergé de plusieurs champs disciplinaires aussi bien en informatique comme l'interaction Homme-machine, l'Informatique Graphique, la visualisation scientifique, les systèmes d'information géographique mais aussi d'autres domaines tels que le design visuel et la psychologie. La visualisation demande l'exécution de trois processus préalables : 1) extraire des données (à partir de différentes sources, notamment les interfaces de programmation) ; 2) les transformer (notamment *via* des méthodes algorithmiques) ; 3) charger ces données transformées (notamment dans un système d'aide à la décision) ; 4) visualiser les données chargées ou traitées (Hachour 2015).

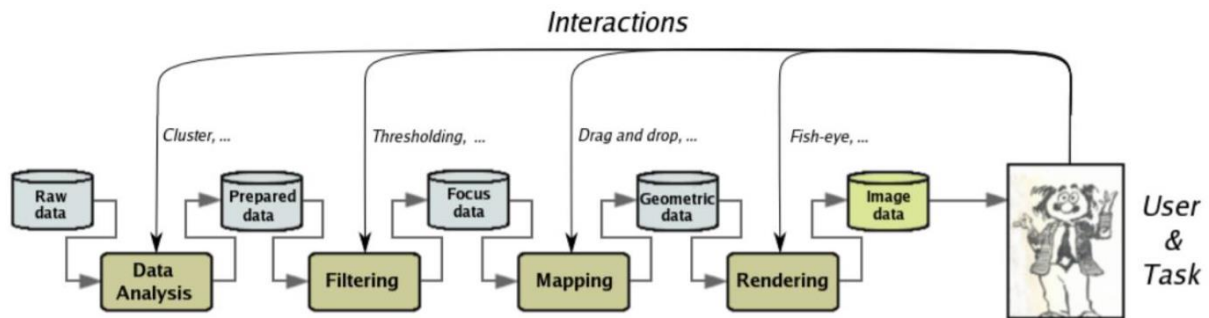
La visualisation des données permet de combiner l'information, d'analyser les tendances et de partager de précieuses données d'affaires de manière conviviale et efficace. Grâce à son aspect visuel, elle rend les importantes quantités d'information plus digeste pour les utilisateurs. Les cartes interactives, les graphiques, les cartes de risques et les infographies sont des exemples de visualisation des données simples et puissants.

Grâce à la visualisation de données, vous pourrez :

- Prendre de meilleures décisions en raison de l'information améliorée.
- Augmenter le nombre d'employés en mesure d'analyser et de mettre à profit l'information existante.
- Avoir une vue d'ensemble sur les informations.

Dans le présent rapport, nous allons expliquer, étape par étape, le processus de visualisation appliqué pour représenter les données du fichiers csv.

2 Processus de génération d'InfoVis



2.1 Analyse de données

La première étape du processus est d'analyser les données qui consiste à nettoyer et filtrer les données, gérer les valeurs manquantes, faire de transformations s'il est nécessaire, normalisation des variables, élimination des doublant ...etc.

Dans notre cas, nous avons subi au fichier de données plusieurs modifications comme :

- Il avait pour chaque colonne deux autres colonnes associé score et commentaire, nous avons constaté que ces deux colonnes sont inutiles. Par conséquent, on a les supprimer.
- Colonne âge, nous avons constaté que la forme de cette colonne n'est pas la même pour toutes les lignes : 30 ans, 25 سنة, 22 an => 30 ans, 25 ans, 22 ans.
- Les valeurs de la colonne numéro 23 « Combien de temps avez-vous passé en position assise ou couchée lors d'une journée habituelle ? » n'étaient pas normalisés : 9h, 6 ساعات, 5 _ 6 h, 5 heures par jour, donc nous avons normaliser la forme.

Nous avons constaté que la question la plus pertinente n'existe pas parmi les colonnes « Souffrez-vous des maladies ? ». Par conséquent, notre étude sera faite à la base des autres questions du formulaire.

2.2 Filtering

Après avoir analyser et nettoyer les données, nous avons classier les attributs selon leur pertinence et leur utilité dans l'étude de la maladie. Nous avons obtenu la classification suivante :

- a. Plus pertinent :
Age, sexe, fumeur ou non, consommation des boissons alcoolique, consommation des fruits, consommation des légumes, pratiquer du sport ou non, existence des maladies dans la famille, nom maladie.
- b. Pertinent :
Niveau d'étude, activité professionnelle, pourquoi vous avez commencé à fumer, matière grasse consommé.
- c. Moins pertinent :
Wilaya, horodateur, souhaiter d'arrêter de fumer.

Nous avons classé wilaya et horodateur les moins pertinents car toutes les réponses étaient dans le même abs du temps, et la moitié des réponses n'ont pas spécifier la wilaya.

Nous avons choisi les cinq variables suivantes pour la représentation visuelle :

1. Sexe
2. Age
3. Fumeur/non-fumeur
4. Pratiquer sport ou non
5. Activité professionnelle.

Après avoir identifier les variables à utiliser, nous avons calculé les statiques associées à ces variables en essayant le maximum de bien mener notre étude sur la maladie.

1) Nombre de fumeur et non-fumeur selon la l'activité professionnelle

Activité professionnelle	Fumeur		Total général
	Non	oui	
Étudiant	66	2	68
Employé(e) de l'Etat	21	1	22
Employé(e) dans le privé	8	3	11
Chômeur(se)	5		5
Total général	100	6	106

Figure 1 Nombre de fumeur et non-fumeur selon la l'activité professionnelle

2) Pratique du sport par sexe

Sports	Sexe
--------	------

	Homme	Femme	Total général
Non	34	28	62
oui	25	19	44
Total général	59	47	106

Figure 2 Pratique du sport par sexe

3) Répartition de la population selon la catégorie professionnelle

Activité professionnelle	Nombre
Étudiant	68
Employé(e) de l'Etat	22
Employé(e) dans le privé	11
Chômeur(se)	5
Total général	106

Figure 3 Répartition de la population par catégorie professionnelle

4) Répartition de la population par sexe

Sexe	Nombres
Homme	59
Femme	47
Total général	106

Figure 4 Répartition de la population par sexe

5) Répartition de la population par tranche d'âge

Tranches d'âge	Nombres
18 - 25	67
24 - 30	16
30 - 35	11
35 - 40	6
40 - 45	2
>45	4
Total général	106

Figure 5 Répartition de la population par tranche d'âge

2.3 Mapping

Après avoir analysé les données et calculer les statistiques, nous avons choisis les variables visuelles pour la représentation des données calculées.

Pour la répartition de la population selon l'âge, le sexe et l'activité professionnelle nous avons la représenter avec un pie chart.

Comme ces variables sont des variables catégorielles, nous avons opté de les représenter sous un diagramme circulaire.

Pour les combinaisons des variables, nous avons opté pour une représentation sous un diagramme de bar et tree diagramme.

Diagramme de bars pour les statistiques fumeur / Non-fumeur par activité professionnelle et les statistiques de fumeur / Non-fumeur par sexe.

Nous avons représenté les statistiques de fumeur / Non-fumeur de selon le sexe et l'âge.

2.4 Rendering

Nous avons représenté les déférentes visualisations avec D3 « Data Driven Document »

Nous avons utilisé des interaction dans le tree diagram.

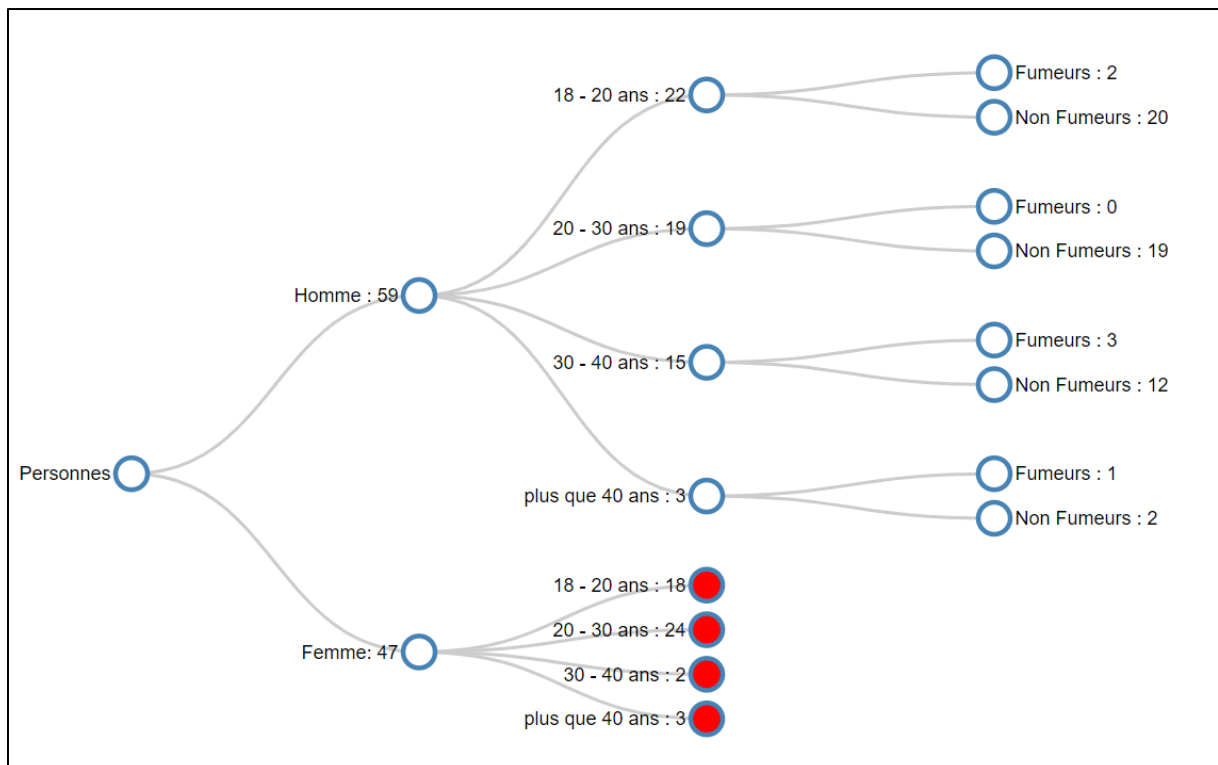


Figure 6 tree diagram représentant les statistiques de fumeur / Non-fumeur de selon le sexe et l'âge

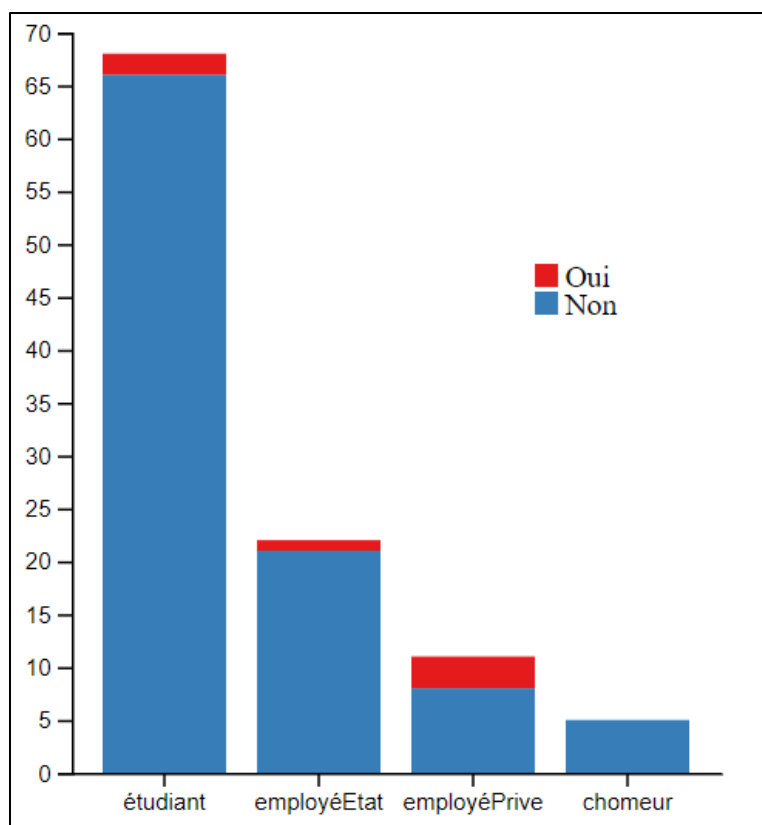


Figure 7 diagramme de bars représentant Statistiques fumeur/Non-fumeur par activité professionnelle

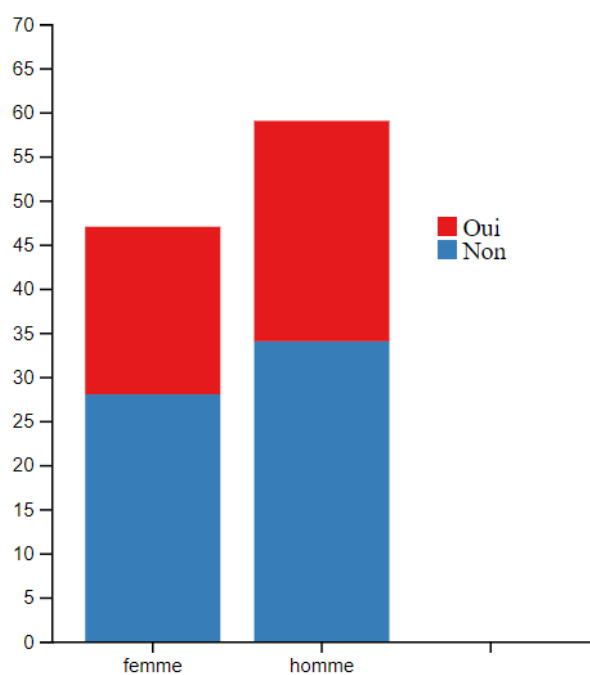


Figure 8 diagramme de bars représentant Statistiques fumeur/Non-fumeur par sexe

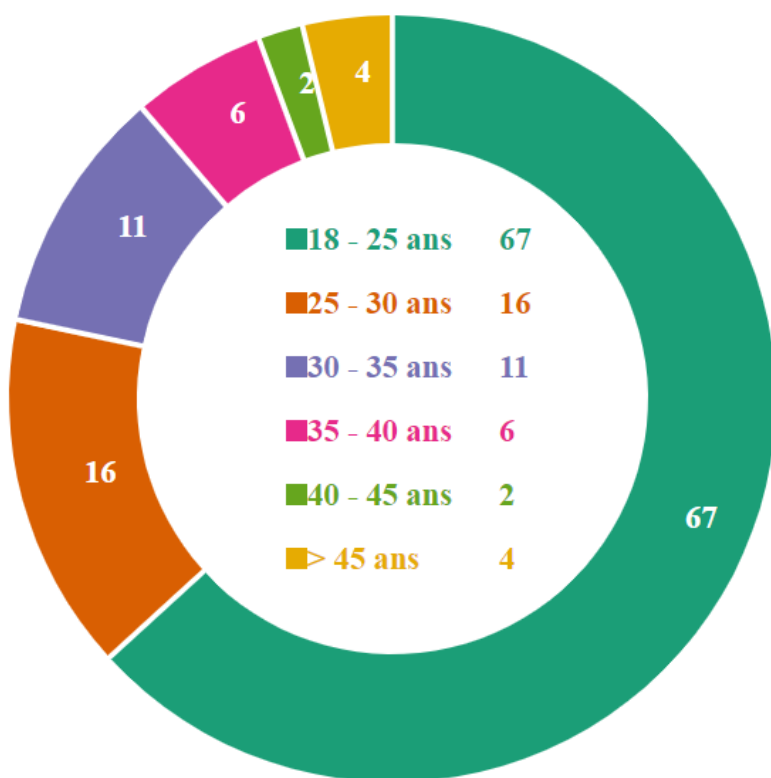


Figure 9 diagramme circulaire représente la répartition de la population par tranche d'âge

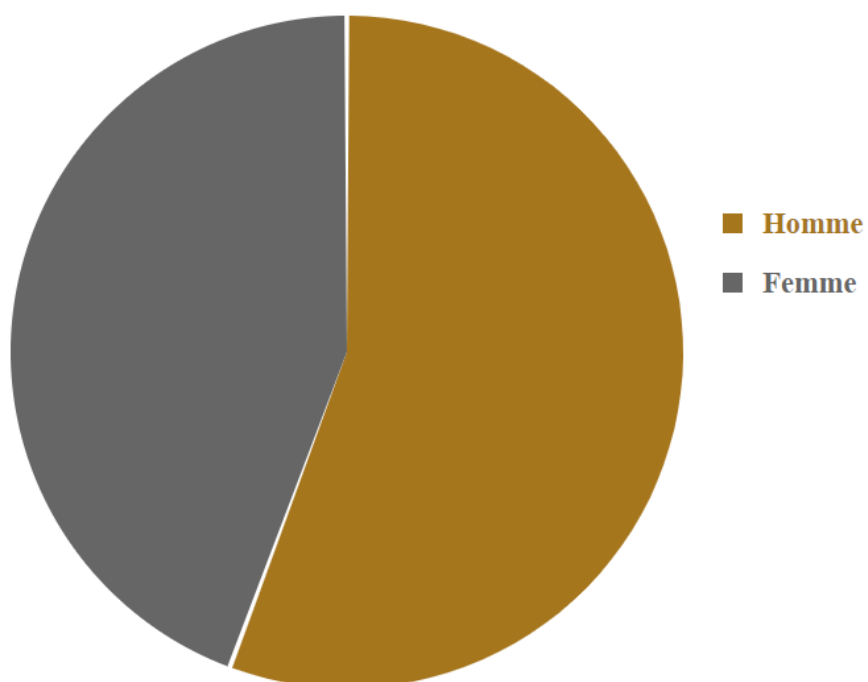


Figure 10 diagramme circulaire représentant la répartition de la population par sexe



Figure 11 diagramme circulaire représentant la répartition de la population par activité professionnelle

3 Conclusion

La visualisation de données ou data vizualisation est la pratique qui consiste à traduire des informations dans un contexte visuel, comme une carte ou un graphique, afin de rendre les données plus faciles à comprendre pour le cerveau humain et d'en tirer des informations. L'objectif principal de la visualisation des données est de faciliter l'identification des modèles, des tendances et des aberrations dans de grands ensembles de données. Le terme est souvent utilisé de manière interchangeable avec d'autres, notamment graphiques d'information, ou graphiques statistiques.

Ce travail personnel et le module nous a permet de savoir identifier les fondements scientifiques d'Infovis, construire des visualisations pour divers problèmes ou données, analyser les modèles d'apprentissages automatiques à l'aide des visualisations, choisir le codage visuel pour chaque type de données et chaque problème associé, lier votre demain de recherche avec le domaine d'infovis et évaluer de différents types de visualisations.

Remarque très importante :

Afin de visualiser les représentation il faut exécuter la commande : `python -m http.server` pour pouvoir récupérer les données depuis les fichiers externes.