

TP 1 : Tableau Public

idir.benouaret@epita.fr

1 Objectifs du TP

- Découvrir Tableau Public
- Avoir une première expérience avec Tableau et construire des visualisations standards avec des jeux de données simples
- Comprendre le mapping de données sur des marques graphiques (ligne, carré, ..) et leurs propriétés (taille, couleur, ..)
- Faire la distinction entre données brutes et abstraites, ainsi qu'entre dimensions et mesures
- Explorer un espace de conception de visualisation
- Pas de distribution pour Linux mais une alternative est la version en ligne : <https://public.tableau.com/app/discover> (il faut créer un compte)

EXERCICE I : Données Covid-19

1. Connecter une source de données :

Télécharger les données de contaminations covid en France (voir moodle : `covid-france.csv`). Lancer Tableau et sélectionner connexion à un fichier texte. Charger ensuite le fichier, et basculer la vue des données (en bas à gauche de Tableau cliquer sur *Source de données (Data Source)*)

Q1 – À quoi correspondent les icônes dans l'en-tête des colonnes ?

Q2 – Lister les principaux types proposés.

Q3 – Quelles colonnes avez vous dû corriger ?

Q4 – Quelles données sont manquantes ?

2. Premiers résultats sous forme de bar chart :

Basculer maintenant sur la 1e Feuille (Sheet 1). Commencer par afficher un graphique avec les jours DAY(Jour) en colonne, et les hospitalisations (Hosp) en ligne. Vous devriez obtenir une courbe des hospitalisations.

Transformer la courbe en diagramme en batons (changer la marque graphique, cf. Figure 1).

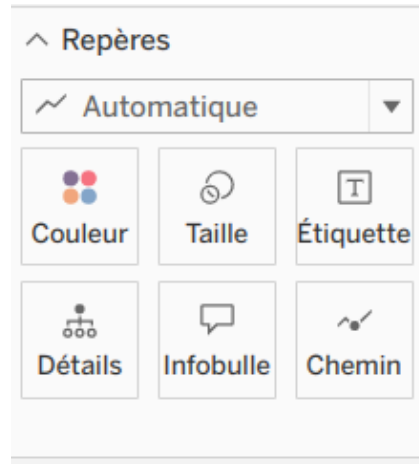


FIGURE 1: différentes options pour les graphiques

Afficher les régions à l'intérieur des barres. Vous remarquerez que Tableau a automatiquement agrégé les données.

Q5 – Dans les données brutes à quoi correspond une ligne d'hospitalisation ?

Q6 – Comment Tableau a-t-il agrégé les données ?

Q7 – Dans quel ordre les régions sont-elles empilées ? Faites en sorte de ré-organiser les régions par ordre décroissant des hospitalisations selon la mesure *Somme*

3. Sauvegarder et partager :

Renommer votre feuille de calcul en double cliquant sur le titre de la visualisation. Sauvegarder, créer un compte Tableau Public si ce n'est pas déjà fait.

4. Filtrer de manière basique :

Pour des raisons géographiques, sociales, démographiques, etc. chacun des DOM devrait être analysé séparément.

Q8 – Filtrer les données de visualisation pour garder uniquement la France métropolitaine

5. Identifier des tendances :

Afficher de nouveau un graphique avec la courbe des hospitalisations, et celles de la positivité (P). Pour ces deux valeurs, les lisser sur 7 jours pour réduire la variabilité due à la remontée irrégulière des données. Pour cela ajouter une "quick table calculation", elle est de deux jours en arrière par défaut, modifier cela pour passer à 7 jours en arrière.

Q9 – Listez les périodes de croissance et de décroissance du virus

Q10 – Quel décalage temporel pouvez-vous identifier visuellement entre les différents indicateurs ?

6. Explorer d'autres représentations :

Sur une nouvelle feuille, explorez d'autres représentations qui vous semblent intéressantes. Essayez de reproduire une visualisation que vous avez pu voir dans la presse ou journal télévisé (heatmaps, effets de la période de vaccination, effets des périodes de confinement, etc.).

EXERCICE II : élections présidentielles (USA)

Télécharger le csv suivant (voir moodle) : `ElectionsPresidentielleUS_data_clean.csv`. Créez un nouveau projet et sélectionner connexion à un fichier texte, charger le fichier, et basculer sur la vue des données (en bas à gauche de Tableau cliquer sur Source de données (Data Source)).

1. Résultats sous forme de table :

Basculer maintenant sur la 1e Feuille (Sheet 1). Commencer par afficher une table avec les années (Year) en colonne, et les États (States) en ligne.

Afficher à l'intérieur de la table le gagnant de l'État en question pour l'année donnée (State Winner). Pour ce faire, associer le State Winner à une marque graphique, ou repère (Mark). Essayer avec du texte (Label) ou de la couleur. Ajuster le format de la marque graphique (texte, carré, ...) Ajuster au besoin la couleur (via le menu de la légende, à droite) pour que le rouge représente les républicains, le bleu les démocrates et l'orange les indépendants.

Un aperçu du résultat attendu est illustré dans la Figure 2

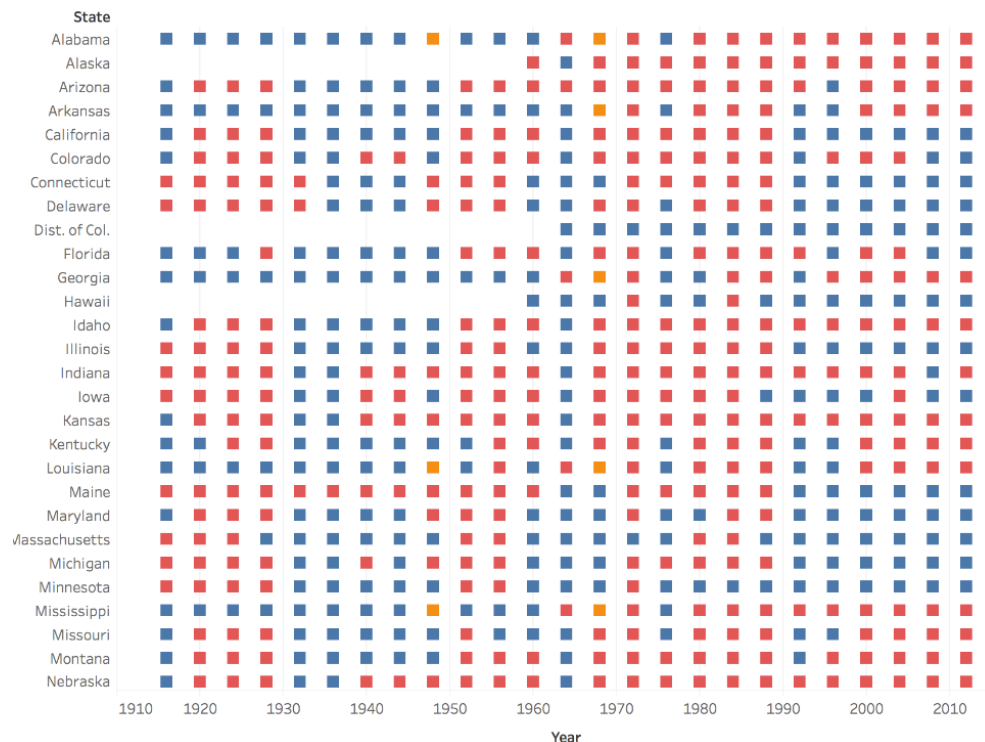


FIGURE 2: graphique des votes aux élections américaines

2. Filtrage basique :

Pour chercher où un Indépendant (candidat non Républicain ou Démocrate) est arrivé en tête, il est possible de cliquer sur la légende (qui est interactive). Cependant créez votre propre mécanisme de filtrage, similaire à la légende, en rajoutant *State Winner* comme filtre. Faire en sorte que le filtre apparaisse à l'écran, et l'utiliser pour répondre à la question suivante.

Q1 – Dans quels États un indépendant est déjà arrivé en tête ?

Q2 – Quelles années un indépendant est arrivé en tête ?

3. Identifier des tendances :

Créer une nouvelle feuille dans votre classeur. Afficher de nouveau une table avec les années (Year) en colonne, et les États (States) en ligne.

Regarder dans les données et identifier quelle colonne peut donner une indication sur la force de la victoire (en bas à gauche de Tableau cliquer sur Source de données (Data Source)). Utiliser une marque textuelle (ou changer celle utilisée par défaut) pour indiquer le vainqueur (State Winner), et la couleur pour indiquer l'intensité de la victoire (avec la variable que vous avez identifié dans les données). Ajuster la couleur pour qu'elles représentent les partis (clic droit sur la légende > modifier les couleurs), et utiliser les paramètres avancés pour que l'échelle soit uniforme. Sauvegarder.

Q3 – Quelles sont les périodes temporelles (intervalles) plutôt démocrates et plutôt républicains.

4. Cartographier :

Créer une nouvelle feuille, ajouter Latitude et Longitude en ligne et colonne

Via le menu *Montre-moi (Show me)* en haut à droite, basculer sur une visualisation sous forme de carte. Rajouter la dimension State comme Repère (Mark) détail, et la variable *RD Difference (bin)* en couleur.

Rajouter les années en Pages. Il est maintenant possible de naviguer entre les années.

5. Exploration des états instables :

Aux États Unis, certains états votent toujours pour le candidat d'un parti donné soit républicain soit démocrate et ceci quelque soit le candidat qui se présente. Mais d'autres états sont plus à même de basculer d'un camps vers un autre. Construire une visualisation qui vous permet d'identifier les États les plus stables politiquement, et/ou ceux qui changent le plus souvent sur les 30 dernières années. Étendre ensuite votre visualisation pour explorer le changement au cours du temps.