Introduction

Table of contents

[Début mise à jour: 03-03-2023]

Par rapport à la première version (2020) et suite aux évolutions qu’a connu le logiciel depuis, il me semble plus pertinent de se focaliser sur certaines points importants concernant la production de graphiques que de les reprendre quasiment un par un selon le types de variables introduites. L’aide de Stata est suffisamment qualitative pour donner les éléments de programmations utiles. On partira plutôt de questions, et voir les réponses les plus adaptés. Contrairement à la première version, on priviliera des solutions programmatives les plus simples, aidés par certaines contributions externes mises à dispositions ces deux dernières années. Citons trois exemples:

* **violinplot** (2022) de Ben Jann pour la visualisation des distributions de variables continues.
* **fabplot** (2021) de Nicolas J.Cox pour la gestion des effets spaghetti.
* **waffle\_plot** (2022) de F.Rios-Avila. On tient peut-être l’alternative au très décrié *camember* (*pie chart*)

**Lien visualisation des données**:

* L’incontournable ***From Data to Viz***: <https://www.data-to-viz.com/> dont il faut lire absolument la partie consacrée aux [mises en garde](https://www.data-to-viz.com/caveats.html)

**Principaux contributeurs**

[**Nicola J.Cox**](https://www.durham.ac.uk/staff/n-j-cox/)  
Très connu par la communauté des utilisateurs du logiciel, NJ.Cox est surement la personne qui le manipule mieux. Il propose régulièrement des notes pédagogiques dans les fameux *stata tips* ou *speaking stata* du *Stata Journal*. Ses commandes sont toujours très utiles, et plus particulièrement pour les graphiques Certaines catplot , tabplot , spineplot , fabplot, axis seront présentées plus loin. Un peu en décalage avec les effets de mode dans le domaine de la visualisation, il privilégie le fond/ le sens plutôt que la forme / l’esthétique…et c’est très bien ainsi. Sa dernière contribution **fabplot** me semble essentielle pour gérer les effets dits *spaghetti*. Egalement amateur de petites phrases un peu piquantes. Par exemple concernant la nécessité de mettre systématiquement un titre aux axes: “*A reader who doesn’t understand what 2005 … 2017 mean won’t understand anything else in the presentation or paper*” (fil de discussion sur Statalist) .

**Quelques visuels de graphiques obtenus avec les commandes de NJ.Cox**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | Fabplot - line | |  | |  | | --- | | Fabplot - scatter | |

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | Catplot | |

[**Ben Jann**](https://github.com/benjann?page=1&tab=repositories) Egalement une référence pour les utilisateurs de Stata, sa contribution dans le domaine des graphiques est majeure. **colorpalette** (chapitre 3) est une commande incontournable pour la gestion des couleurs et, personnellement, je ne peux plus me passer de **grstyle**. Très recemment la commande **violinplot** donne une solution complètement aboutie pour la visualisation d’une distribution. Toujours dans les ***must have***, **heatplot**/**hexplot** reste un incontournable pour la visualisation des distributions jointes. B.Jann reste également particulièrement connu pour sa suite **coefplot**.

**Quelques visuels de graphiques obtenus avec les commandes de Ben Jann**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | [Violinplot](https://github.com/benjann/violinplot/blob/main/README.md) | |  | |  | | --- | | Hexplot | |

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | Heatplot | |

[**Asjad Naqvi**](https://github.com/asjadnaqvi) A fait une entrée assez fracassante dans le domaine de la visualisation avec Stata en 2020. Il est l’auteur d’une douzaine de commandes sous l’angle de la fameuse ***dataviz*** moderne. Compte tenu du moteur vieillissant de Stata, le tour de force est particulièrement impressionnant. Néanmoins, de mon point de vue, une partie de ces visualisations prennent tout leur sens de manière intéractive, avec une diffusion type html. Cela demande alors de se reporter vers d’autres langage, comme **python**, maintenant directement exécutable dans l’interface principale de Stata. On est donc ici dans une approche plus moderne, mais également complexe de la visualisation.  
Sur son site du réseau [medium](https://asjadnaqvi.medium.com/), les solutions graphiques proposées sont toujours précédées d’une partie très détaillée sur la mise en forme des données. Une petite remarque sur *medium*: sans inscription au site, le nombre d’accès fait l’objet d’un nombre de pages limité mensuellement. Cela est fort dommage. Penser donc à récupérer les programmes qui sont mis à disposition par l’auteur.

**Quelques visuels des commandes d’A.Naqvi**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | [Tree map](https://github.com/asjadnaqvi/stata-treemap) | |  | |  | | --- | | [Arc plot](https://github.com/asjadnaqvi/stata-arcplot) | |

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | [Stream plot](https://github.com/asjadnaqvi/stata-streamplot) | |

[**Demetris Christodoulou**](https://graphworkflow.com) A ma connaissance il n’a pas programmé de commande, mais la démarche de son support, très orienté sémiologie graphique est très pertinente, ses choix graphiques faisant l’objet d’une justification. Les programmes sont mis à disposition, et sa conversion en Stata du célèbre graphique de *Charles Joseph Minard* ([original](https://fr.m.wikipedia.org/wiki/Fichier:Minard.png)) sur la campagne de Russie des armées napoléonienne est vraiment bluffante. Dans les prévisions de publication de Stata, il serait en train de préparer un ouvrage pour 2024 (*Visual Data Analytics*)

**Quelques graphiques de D.Christodoulou**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | [Minard](https://graphworkflow.com/2019/06/25/minard/) | |  | |  | | --- | | [Massacre dauphins](https://graphworkflow.com/2018/12/16/taiji-cove/) | |

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | [Mariage homo](https://graphworkflow.com/2020/01/17/marriage-equality/) | |

De manière générale, ne pas hésiter à aller également sur le forum [statalist](https://www.statalist.org/) où les questions concernant les graphiques sont relativement nombreuses, et les réponses apportées rapides et pertinentes.