
CONCEPTS AVANCÉS DE LA VISUALISATION DES DONNÉES ET DE LA CRÉATION DE RAPPORTS

PRÉPARATION DU TERRAIN



APERÇU

1. Visualisation des données avec ggplot2
2. Introduction aux tableaux de bord

GRAPHIQUES EN LANGAGE R

En 2018, il existe quatre fonctions graphiques offertes en langage R, soit :

- *base*
- *grid*
- *lattice*
- *ggplot2*

L'accès varie avec la fonction: les fonctions *base*, *grid* et *lattice* sont incluses dans l'installation de base; les fonctions *grid*, *lattice* et *ggplot2* doivent être chargées avant leur utilisation.

INTRODUCTION À GGLOT2

La fonction *ggplot2* est en fait un jeu d'outils permettant de transformer des données en éléments d'affichage visuels. Elle permet à l'utilisateur de commander les détails de l'affichage graphique.

Aspect le plus important, la fonction *ggplot2* peut servir à établir la **structure logique** du graphique.

Un graphique *ggplot2* comporte deux éléments principaux (et des termes optionnels) :

- une fonction esthétique (**aes** – liens entre les données et les éléments graphiques)
- une fonction de géométrie (**geom** – type de graphique)
- *facets, *coordinates, *scales, *labels, *guides, etc.

GRAMMAIRE DE GGLOT2

1. Tidy Data

```
p <- ggplot(data = gapminder, ...
```

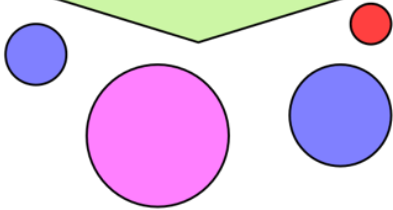
gdp	lifexp	pop	continent
340	65	31	Euro
227	51	200	Amer
909	81	80	Euro
126	40	20	Asia

2. Mapping

```
p <- ggplot(data = gapminder, mapping =  
  aes(x = gdp, y = lifexp, size = pop,  
      color = continent))
```

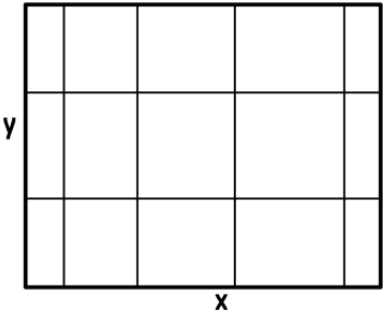
3. Geom

```
p + geom_point()
```



4. Co-Ordinates & Scales

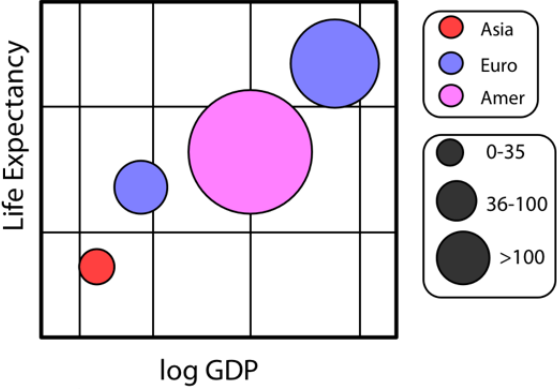
```
p + coord_cartesian() + scale_x_log10()
```



5. Labels & Guides

```
p + labs(x = "log GDP", y = "Life  
Expectancy", title = "A Gapminder Plot")
```

A Gapminder Plot



GRAMMAIRE DE GGLOT2 – GEOM

La source de données et les variables sont précisées au moyen de `ggplot()`.

Les diverses fonctions `geom` précisent la **manière** dont ces variables seront représentées visuellement :

- au moyen de points, de barres, de lignes, de zones ombragées, etc.

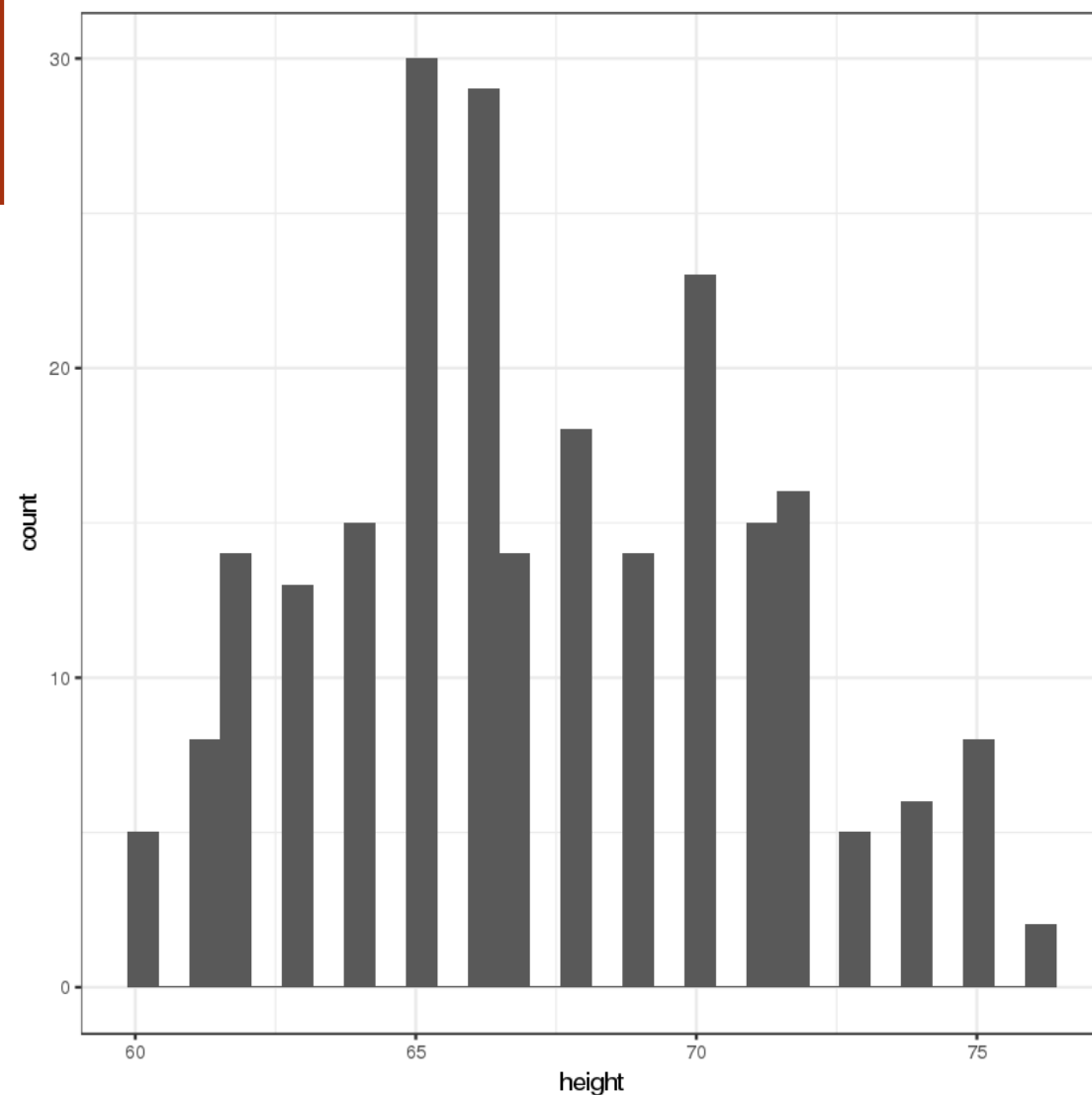
Il existe présentement 37 géométries.

GRAMMAIRE DE GGLOT2 – GEOM()

```
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
  geom_histogram()

# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
  geom_boxplot()
```

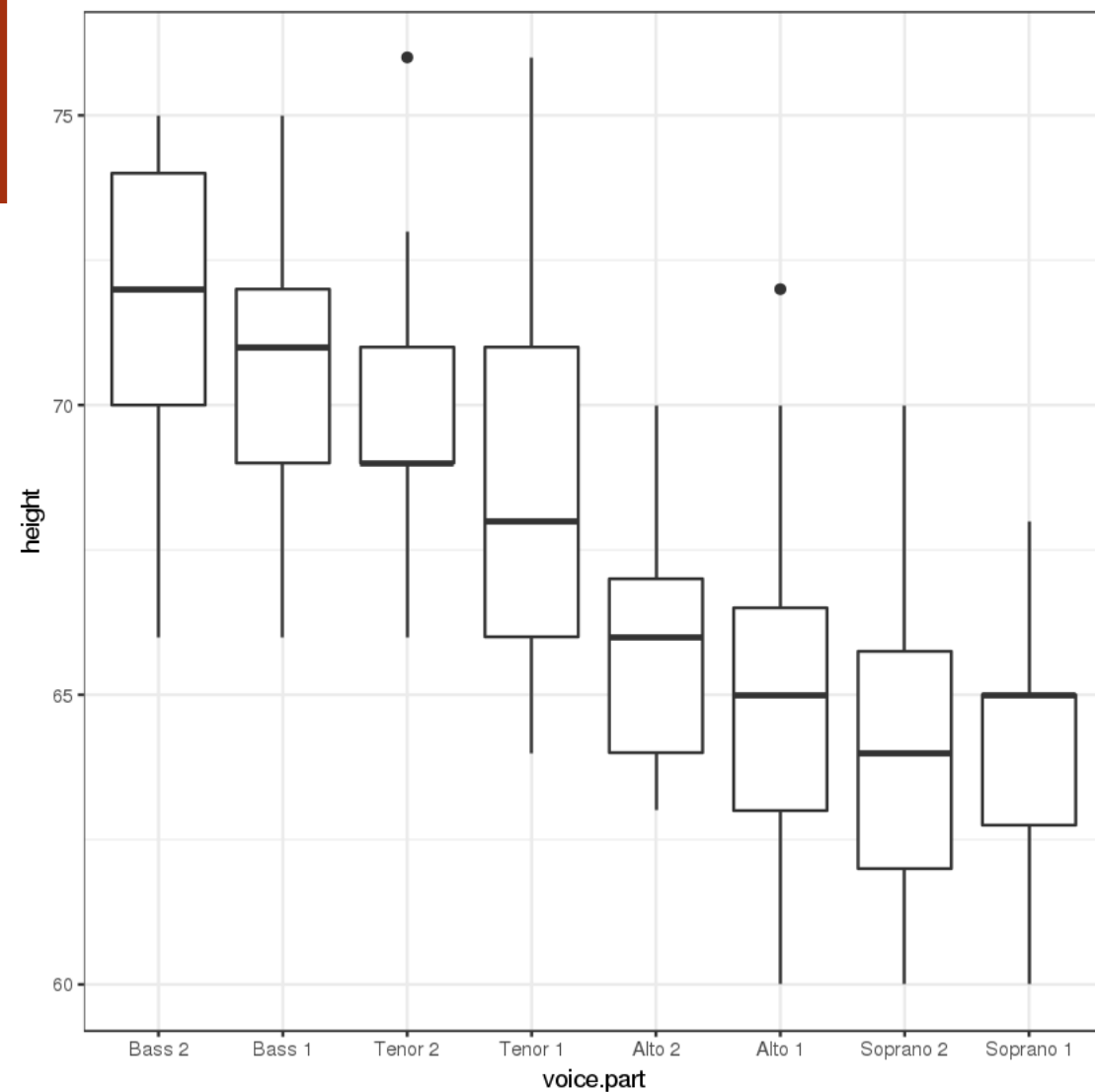


GRAMMAIRE DE GGLOT2 – GEOM()

```
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
  geom_histogram()

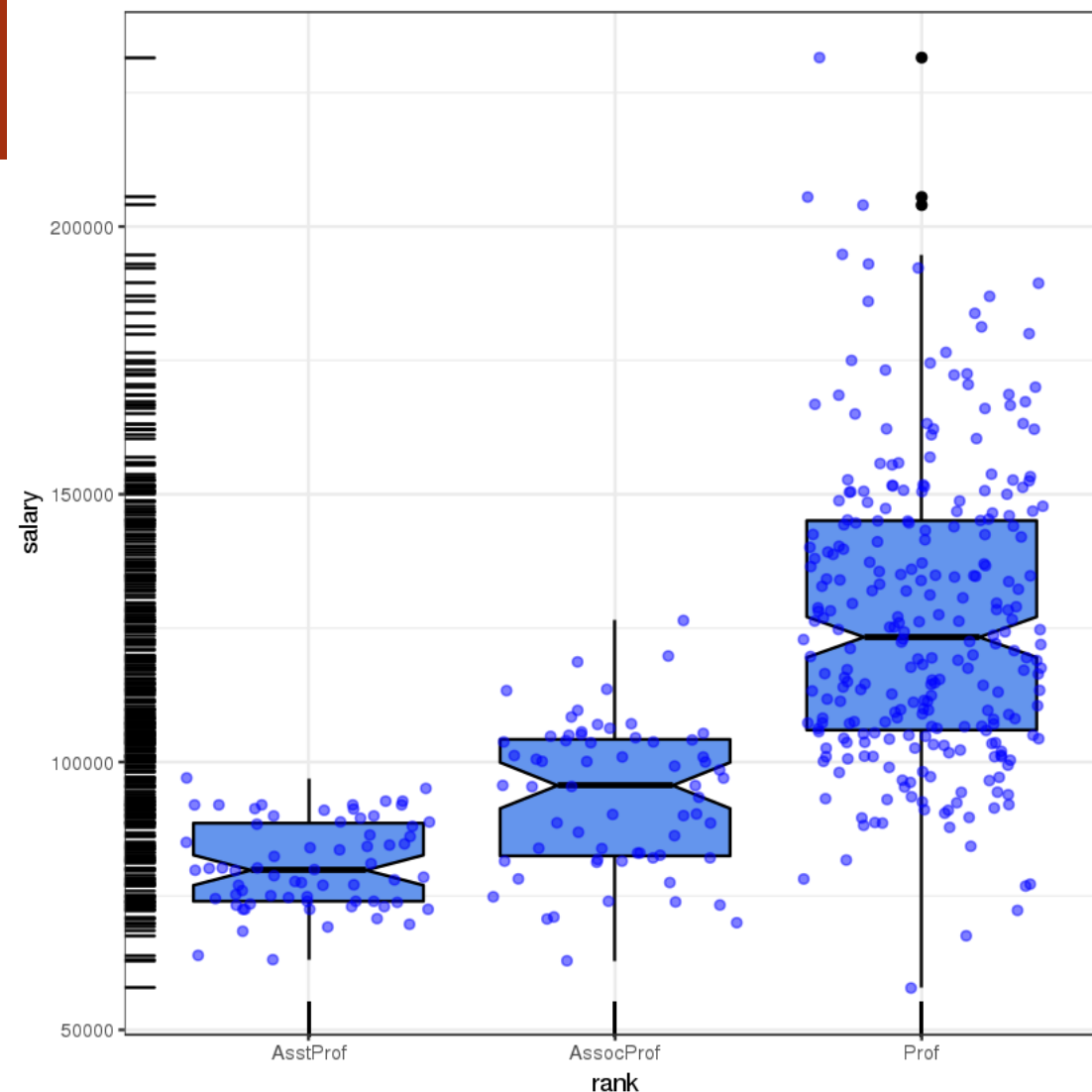
# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
  geom_boxplot()
```



GRAMMAIRE DE GGLOT2 – GEOM()

```
library(ggplot2)
data(Salaries, package="car")
# Using data on salaries of a sample of
# US university professors (2018-2019)
# var: rank, sex, yrs.since.phd, yrs.service, salary

ggplot(Salaries, aes(x=rank, y=salary)) +
  geom_boxplot(fill="cornflowerblue", color="black", notch=TRUE) +
  geom_point(position="jitter", color="blue", alpha=.5) +
  geom_rug(side="l", color="black")
```



GRAMMAIRE DE GGLOT2 – ESTHÉTIQUE

L'**esthétique** désigne les attributs affichés des données.

Vous devez faire correspondre une donnée à un attribut (comme la taille ou la forme d'un repère) et créer la légende appropriée.

Vous précisez l'esthétique au moyen de la fonction `aes ()`.

Vous pouvez préciser l'esthétique dans la fonction *data* ou *geom*. Si vous précisez l'esthétique dans la fonction *data*, l'esthétique vise alors toutes les fonctions *geom* précisées.

GRAMMAIRE DE GGLOT2 – ESTHÉTIQUE

Les caractéristiques esthétiques offertes avec `geom_point()` (scatterplot), p. ex., sont les suivantes :

- `x, y, alpha, color, fill, shape, size`

Il existe une **différence importante** entre les caractéristiques (comme la couleur et la forme) selon qu'elles sont précisées à l'intérieur et à l'extérieur de la fonction `aes()` :

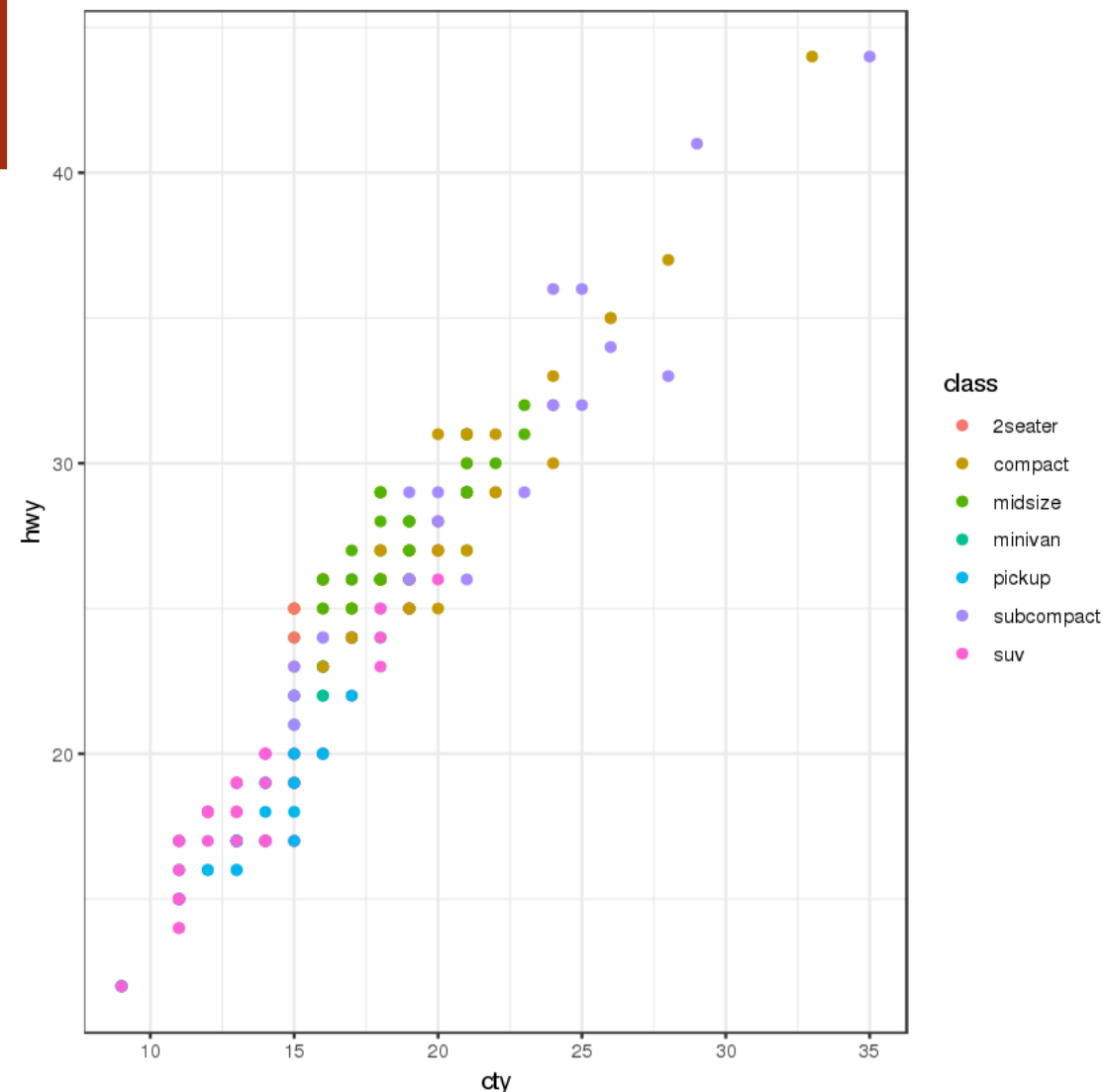
- à l'intérieur : la couleur ou la forme choisie repose automatiquement sur les données
- à l'extérieur : la caractéristique ne vise pas les données.

GRAMMAIRE DE GGLOT2 – AES()

```
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(colour = "red")
```

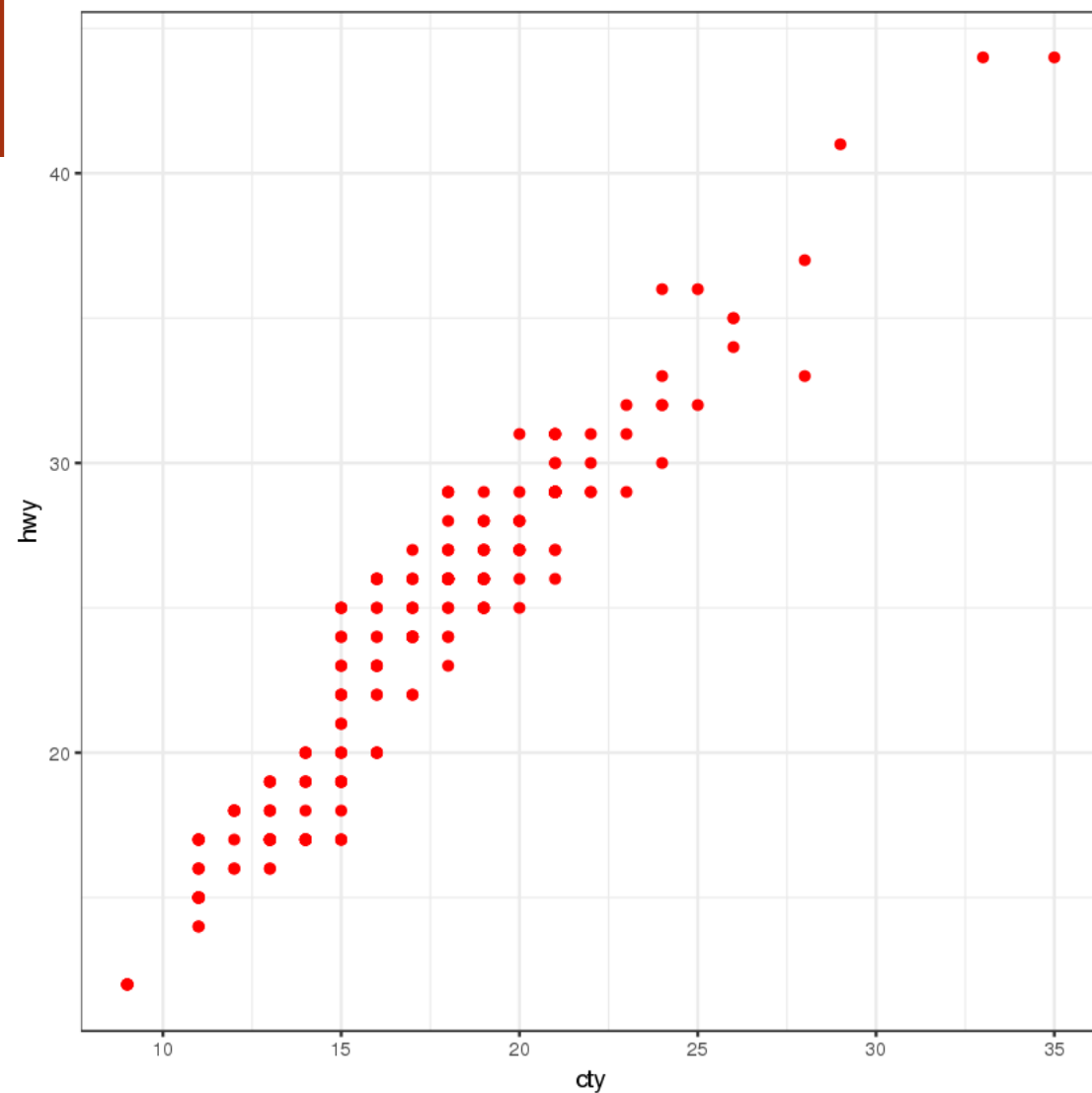


GRAMMAIRE DE GGLOT2 – AES()

```
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(colour = "red")
```



GRAMMAIRE DE GGLOT2 – FACETTES

Avec *ggplot2*, de petits multiples sont désignés des **facettes** :

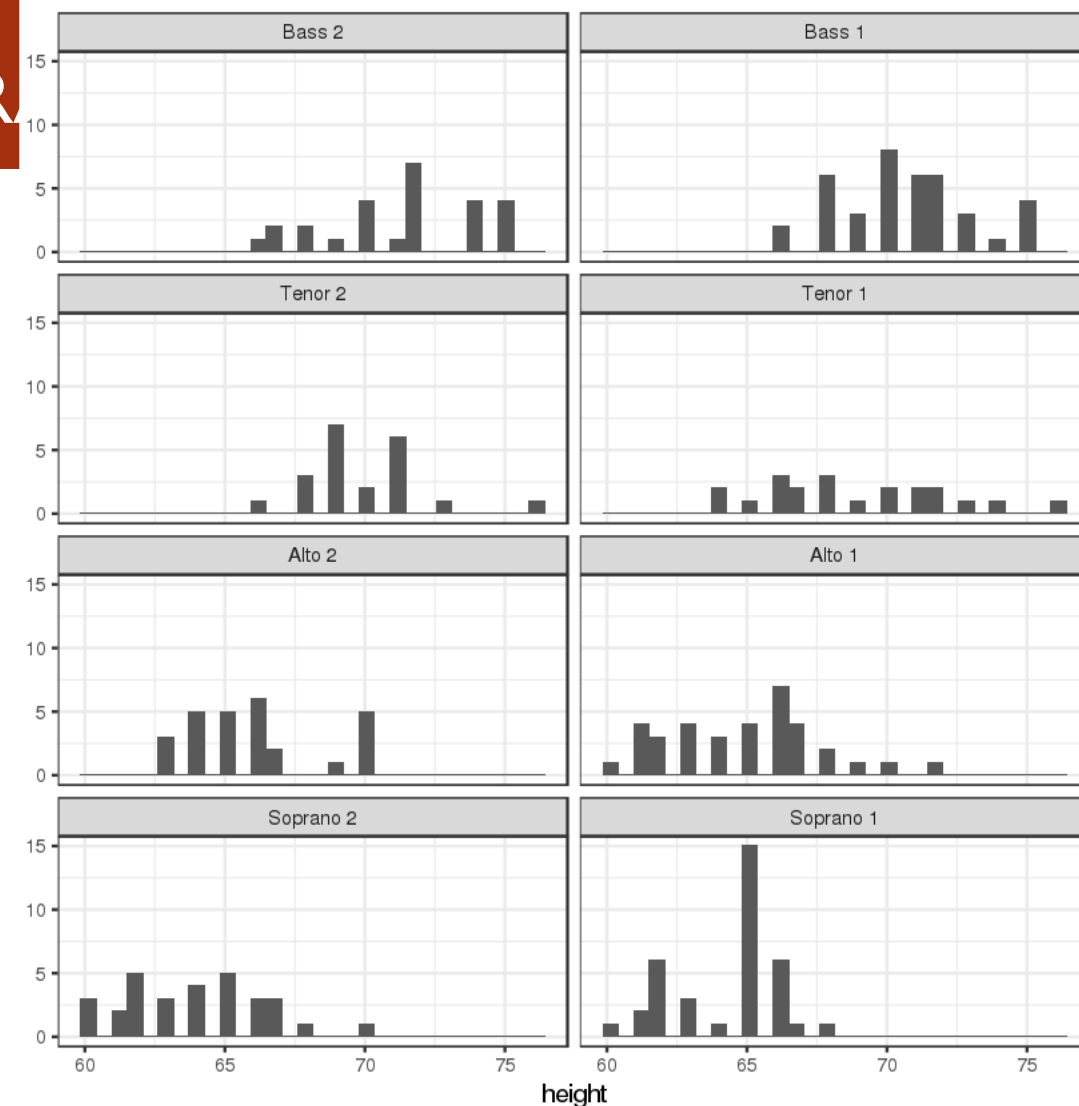
- `facet_wrap()`, `facet_grid()`

Par défaut, tous les graphiques (un pour chaque facteur) partagent les mêmes axes (*scale-wise*).

La séparation du graphique en une suite de graphiques plus petits et disposés côte à côte facilite les comparaisons.

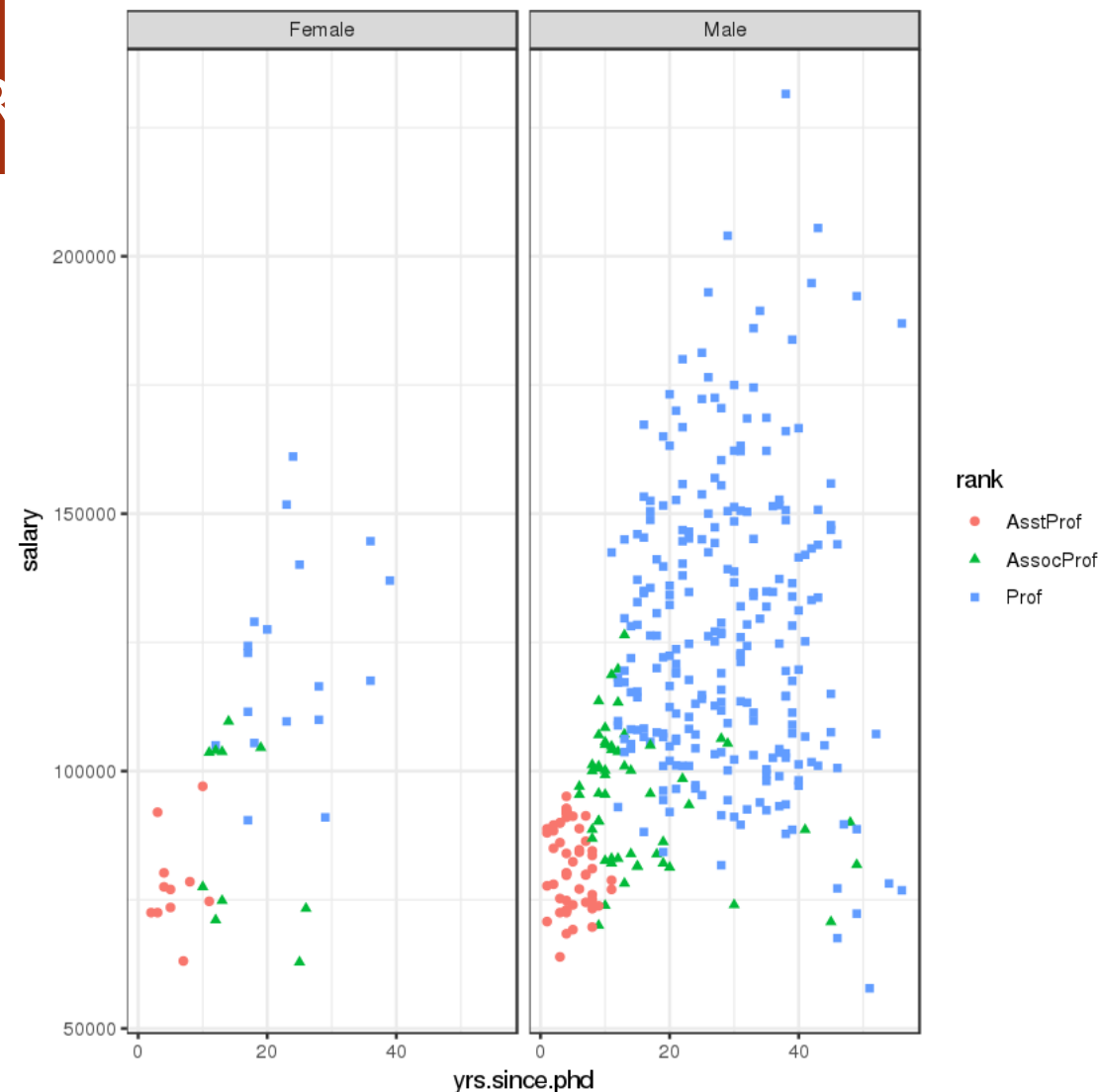
GRAMMAIRE DE GGLOT2 – FACET_WRAP

```
data(singer, package="lattice")  
library(ggplot2)  
ggplot(data=singer, aes(x=height)) +  
  geom_histogram() +  
  facet_wrap(~voice.part, nrow=4)
```



GRAMMAIRE DE GGLOT2 – FACET_GRID

```
data(Salaries, package="car")
library(ggplot2)
ggplot(Salaries, aes(x=yrs.since.phd,
  y=salary, color=rank, shape=rank)) +
  geom_point() +
  facet_grid(~sex)
```



RAPPORT ET DÉPLOIEMENT

Une analyse est bonne seulement si on la **transmet** ou si on la **déploie**.

Questions essentielles :

- Qui reçoit le rapport?
- Quels sont les flux de travail menant à sa création?
- Est-ce que les données peuvent donner lieu à des politiques utiles?

La création automatique de rapports devrait **régulièrement** faire l'objet d'une vérification et d'une validation.

RAPPORT ET DÉPLOIEMENT

La **communication** devrait avoir lieu à diverses étapes du projet, et non seulement à son achèvement :

- vous devez maintenir les commanditaires et les clients au fait des principaux points;
- vous pouvez délaissé les détails techniques, mais vous devez tout de même les documenter.

Le **scénario idéal** consiste à utiliser un logiciel d'analyse qui permet aussi de créer des rapports :

- minimise l'erreur humaine liée à la fonction copier-coller;
- supprime le besoin de maintenir la séparation entre l'analyse et la création de rapports;
- facilite le partage du travail avec les autres membres du projet.

Vous pouvez simplifier davantage le processus en procédant à un déploiement directement sur le web.

TABLEAUX DE BORD

Un **tableau de bord** est un affichage visuel des données qui sert à surveiller les états et à faciliter la compréhension.

Exemples :

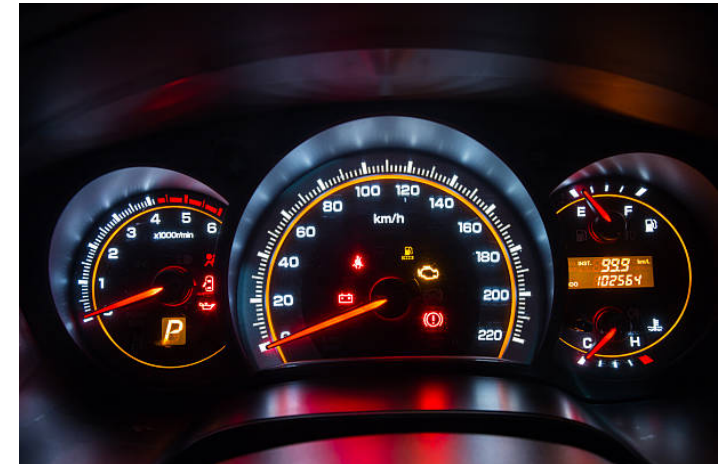
- affichage interactif qui permet à l'utilisateur d'explorer les réclamations d'assurance automobile selon la ville, la province, l'âge du conducteur, etc.;
- fichier PDF qui montre les principaux paramètres de vérification et qui est envoyé par courriel chaque semaine au DG d'un ministère;
- écran monté au mur qui montre en temps réel les statistiques d'un centre d'appel;
- application mobile qui permet aux administrateurs d'un hôpital de voir les délais d'attente chaque heure et chaque jour pour l'année courante et l'année précédente.

QUELQUES QUESTIONS À PRENDRE EN COMPTE

Sur le tableau de bord d'une automobile, l'automobiliste doit comprendre **d'un coup d'œil** un petit nombre d'**indicateurs importants** (vitesse, niveau d'essence, phares, etc.). Un tableau de bord qui ne tient pas compte de ces deux caractéristiques peut donner lieu à des conséquences catastrophiques.

Vous devez répondre aux questions suivantes avant de concevoir un tableau de bord :

- Qui est l'**utilisateur** du tableau de bord?
- Quels renseignements doit **transmettre** le tableau de bord?
- Quelles données (catégories) seront utilisées?
- Qu'est-ce qui **figurera** dans le tableau de bord?
- Comment le tableau de bord va-t-il **aider** l'utilisateur?



LIGNES DIRECTRICES LIÉES À LA CONCEPTION D'UN TABLEAU DE BORD

Nick Smith propose les six règles d'or suivantes :

- **Tenez compte du public** (Qui voulez-vous informer? Est-ce que le DG a réellement besoin de savoir que les serveurs fonctionnent à 88 p. cent de leur capacité?)
- **Sélectionnez le bon type de tableau de bord** (opérationnel, stratégique, analytique)
- **Groupez les données logiquement, utilisez intelligemment l'espace** (séparez les secteurs fonctionnels : produit, ventes et marketing, finances, personnes, etc.)
- **Utilisez des données adaptées au public** (portée et étendue des données, différents tableaux de bord pour différents services, etc.)
- **Évitez d'encombrer le tableau de bord** (présentez seulement les paramètres les plus importants)
- **Actualisez les données à la fréquence appropriée** (en temps réel, chaque jour, chaque semaine, chaque mois, etc.)



TRANSPORTATION



LIVABILITY



ENVIRONMENT



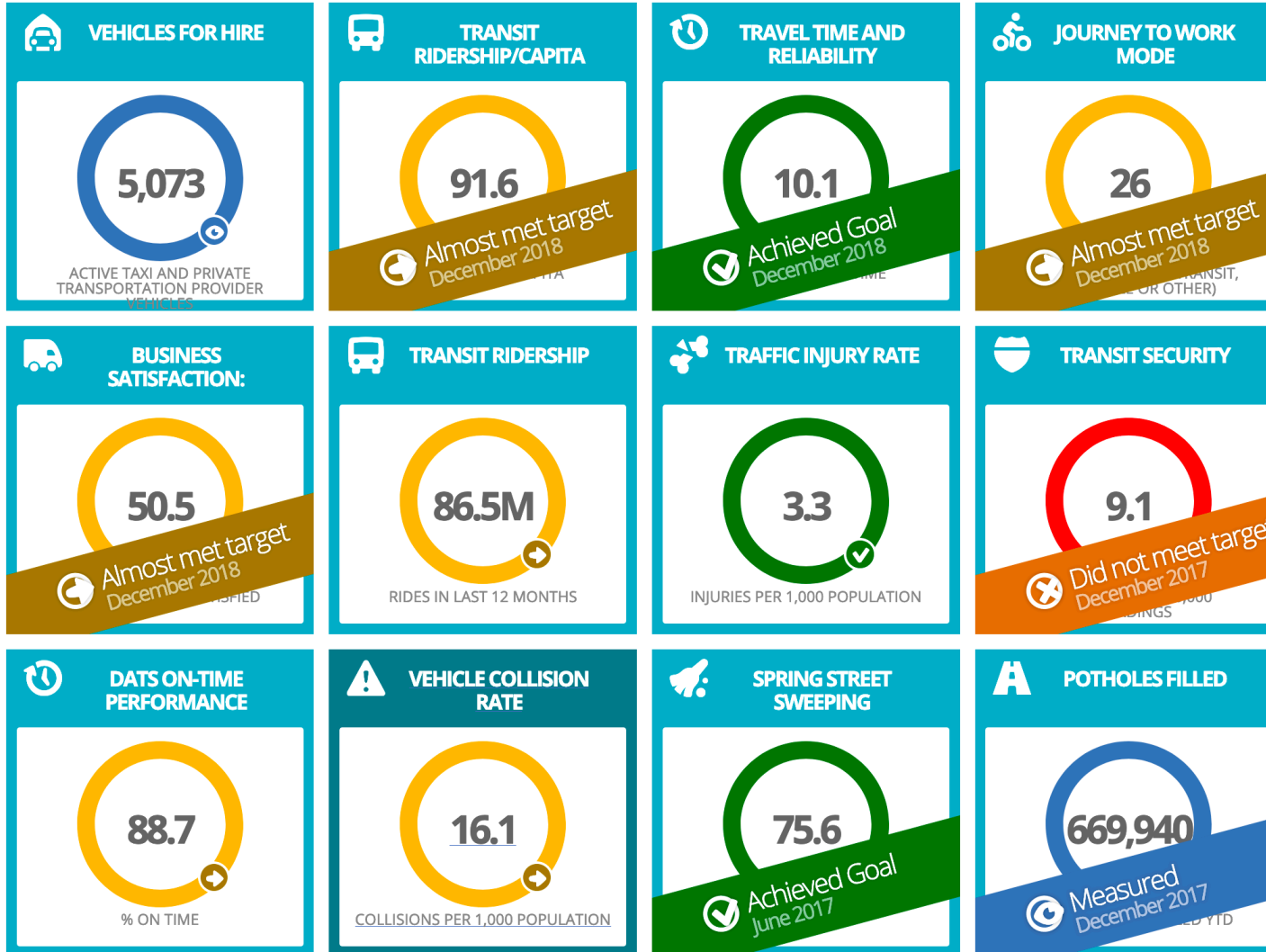
URBAN FORM



ECONOMY



FINANCE

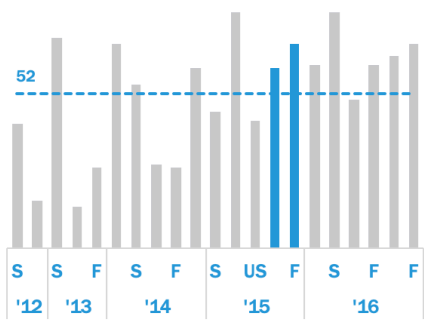


✔ Meets or Exceeds Target ➕ Near Target ✖ Needs Improvement ⚙ Measuring 📊 Collecting Data

Course Metrics

[<https://bigbookofdashboards.com/dashboards.html>]

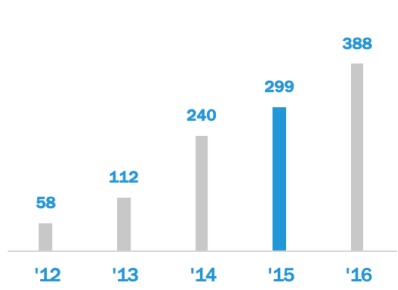
Students



1097

Total Students in five years

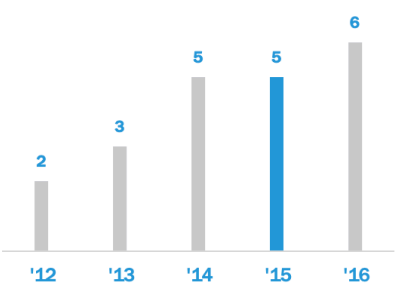
Enrollments



687

Total Students in 2015-2016

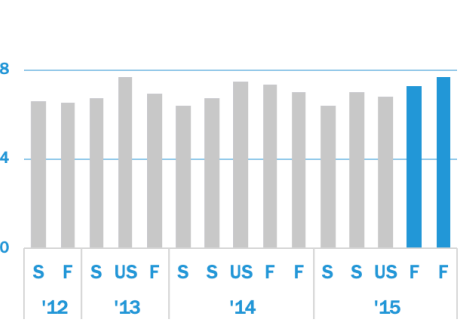
Classes



21

Total Classes in five years

Ratings



7.7 of 8

Most recent instructor rating (out of 8.0)

Semesters

2015 Fall Semester 001	I developed specific skills and competencies
	Overall, this was an excellent course
	The instructor communicated clearly
	The Instructor graded fairly
	The instructor was well organized
	The instructor interacted well with students
2015 Fall Semester 002	Overall, this instructor was excellent
	I developed specific skills and competencies
	Overall, this was an excellent course
	The instructor communicated clearly
	The Instructor graded fairly
	The instructor was well organized
	The instructor interacted well with students
	Overall, this instructor was excellent

Questions

I developed specific skills and competencies
Overall, this was an excellent course
The instructor communicated clearly
The Instructor graded fairly
The instructor was well organized
The instructor interacted well with students
Overall, this instructor was excellent
I developed specific skills and competencies
Overall, this was an excellent course
The instructor communicated clearly
The Instructor graded fairly
The instructor was well organized
The instructor interacted well with students
Overall, this instructor was excellent

● BANA ■ College ● Shaffer

Ratings

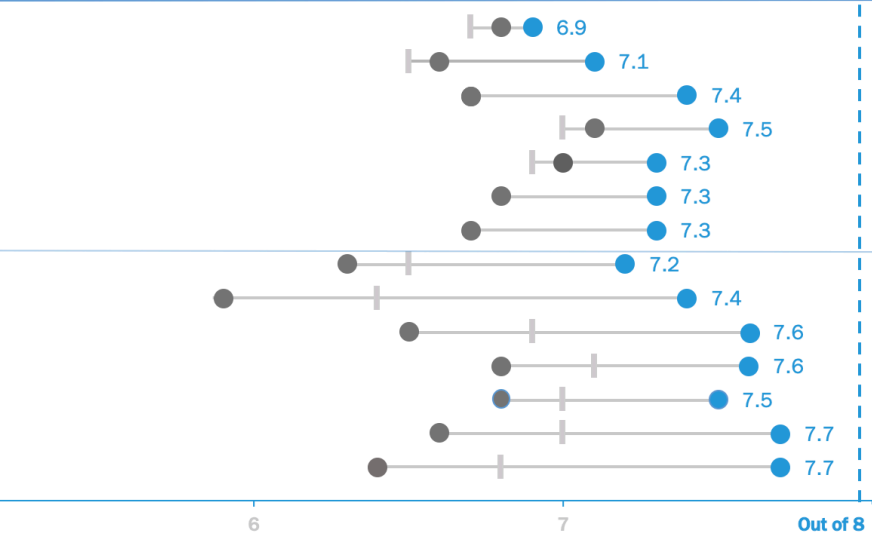


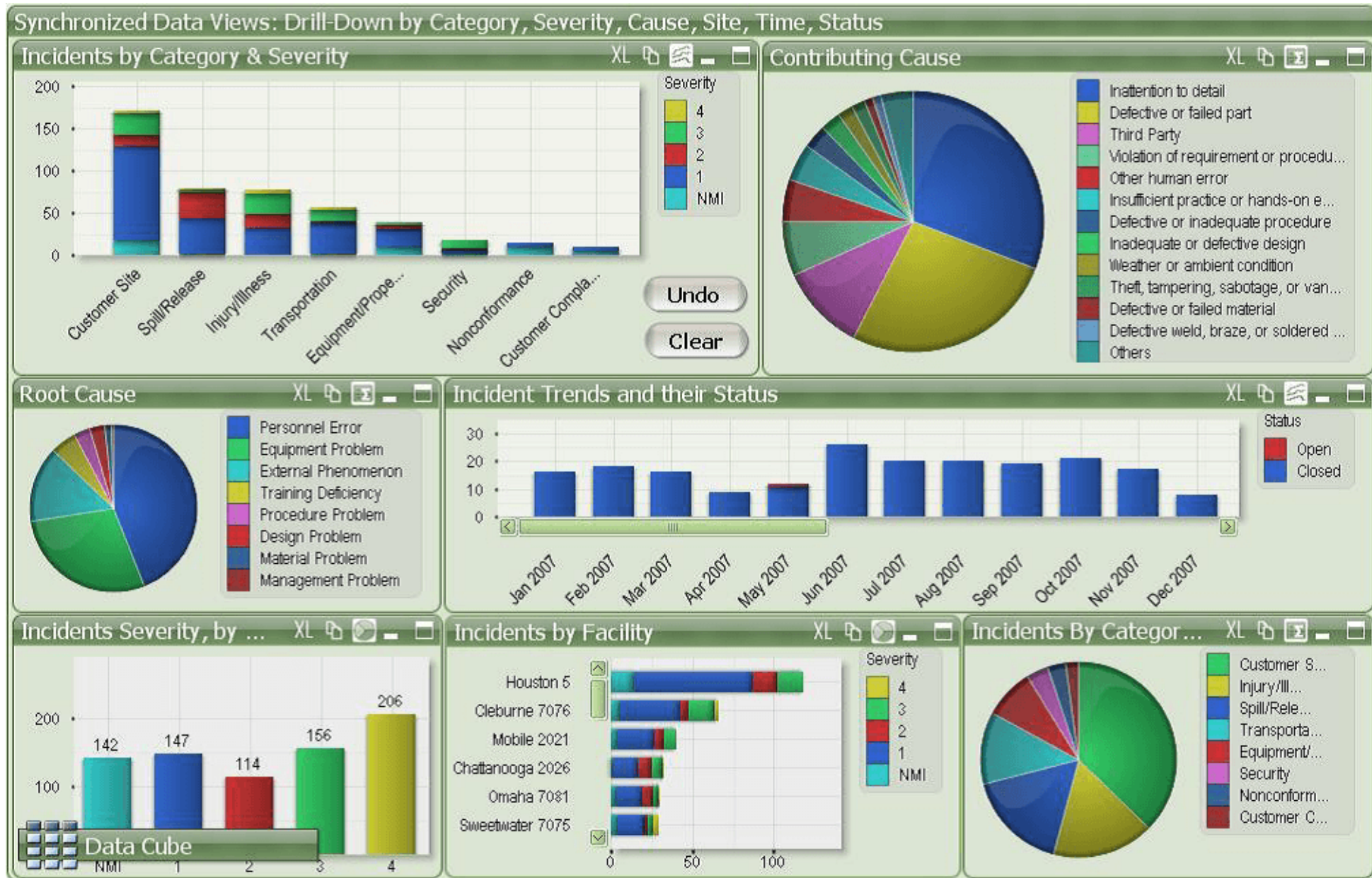
TABLEAU DE BORD – POINTS FORTS

Principaux paramètres faciles à voir

Palette de couleurs simplifiée

Possibilité d'un tableau statique ou interactif

Clarté du sommaire et des détails



What-If Analysis: Impact of Minimum Wage

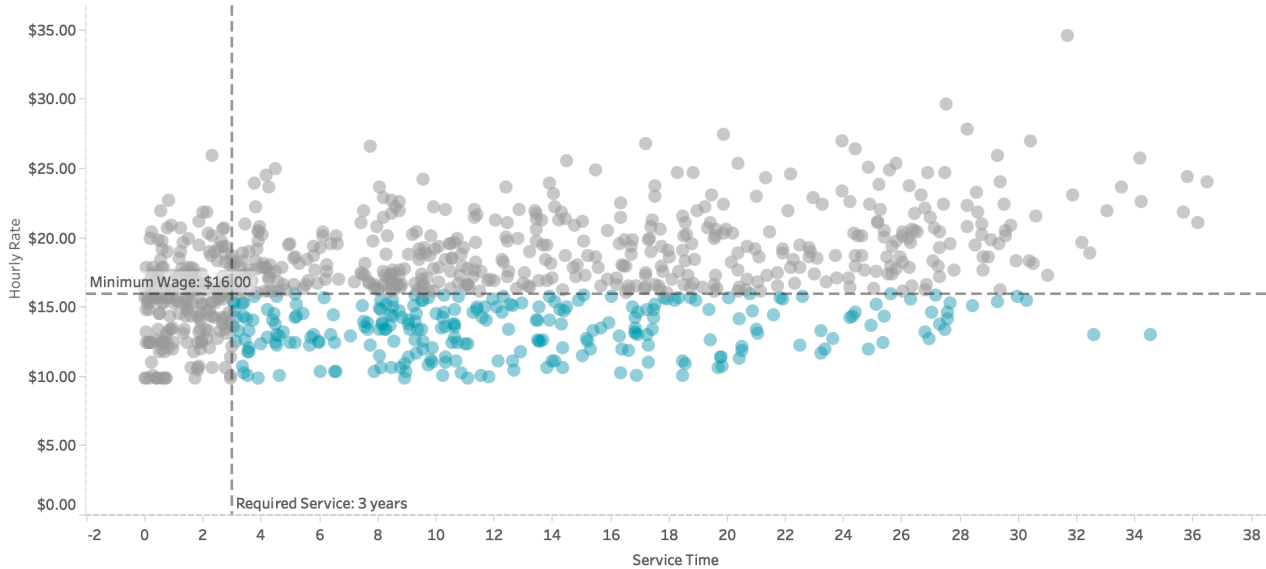
[<https://bigbookofdashboards.com/dashboards.html>]



Proposed Minimum Wage

Required Service

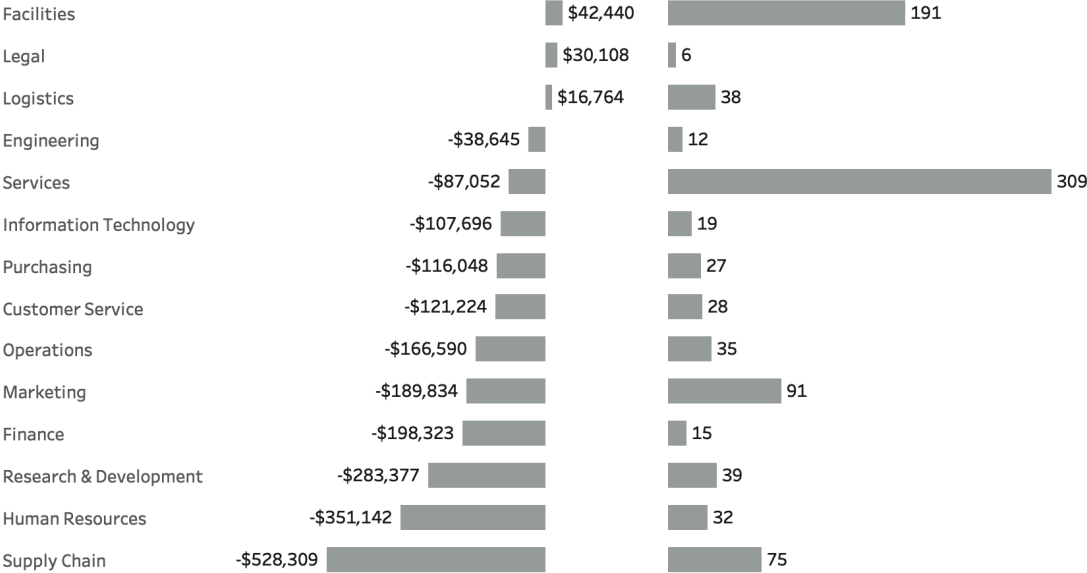
Developed by Matt Chambers
<http://sirvialot.blogspot.com/>



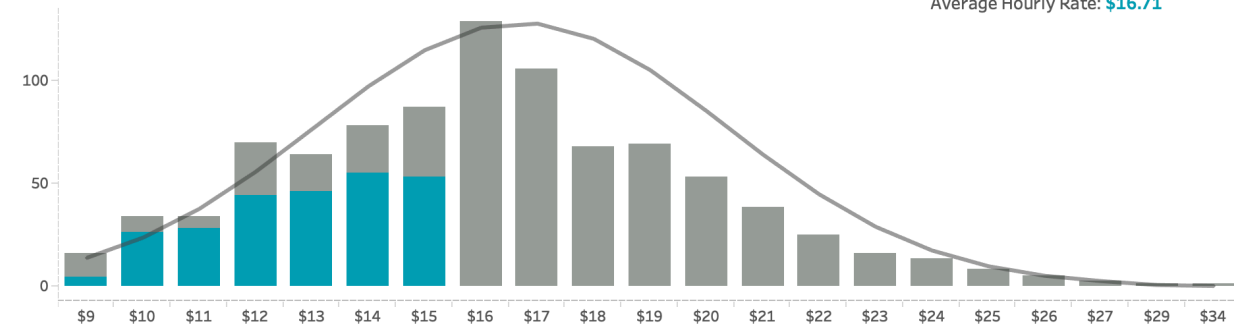
Dollar Impact of Minimum Wage: **\$1,792,206**

Employees Below Minimum Wage: **661**

Employees Below Minimum Wage: **256**



Current Distribution



Distribution with Minimum Wage

