

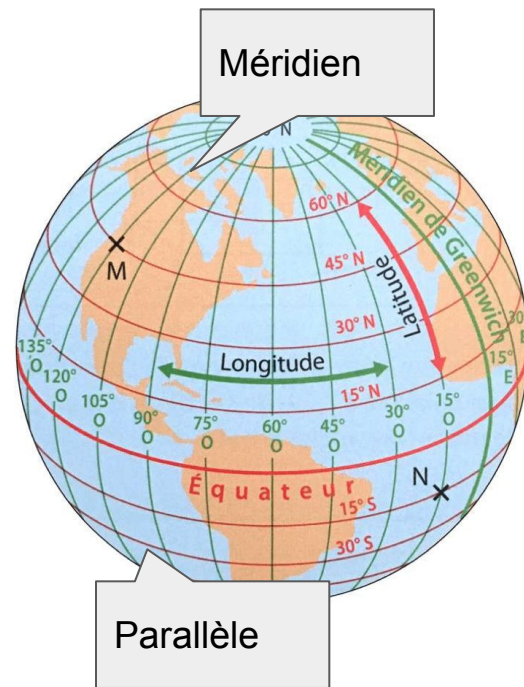
Programmation Statistique 2

Les données géographiques et R

Davide Buscaldi

Quelques notions de base en cartographie

- Comment localiser un lieu sur Terre?
- La Terre est une sphère (3 dimensions)
- Un lieu est défini par des coordonnées géographiques:
 - la **latitude** (*lat*): distance angulaire entre l'équateur et le lieu
 - la **longitude** (*lon*): distance angulaire entre le méridien de Greenwich et le lieu



Coordonnées

- Les coordonnées peuvent être exprimés sous différents formats :
 - degrés **sexa-décimaux** (heures, minutes, secondes)
 - $40^{\circ} 26' 46''$ N $79^{\circ} 58' 56''$ W
 - degrés et **minutes décimaux**:
 - 40° **26.767'** N 79° **58.933'** W
 - **degrés décimaux**:
 - +40.446 -79.982
 - latitude positive = N, négative = S
 - longitude négative = W, longitude positive = E
- Certains systèmes utilisent des coordonnées métriques:
 - ex. *Lambert93*: X : 923825 Y : 6307755

Coordonnées - conversions

- Il y a 60 minutes dans un degré, et 60 secondes dans une minute
- Donc pour convertir du format DMS aux décimaux:
 - $\text{décimaux} = \text{degrés} + \text{minutes}/60 + \text{secondes}/3600$
- Inversement:
 - $D = \text{degrés}; M=60 * (\text{degrés décimaux} - \text{degrés}); S= 3600 * (\text{degrés décimaux} - \text{degrés}) - 60 * \text{minutes}$

Projections

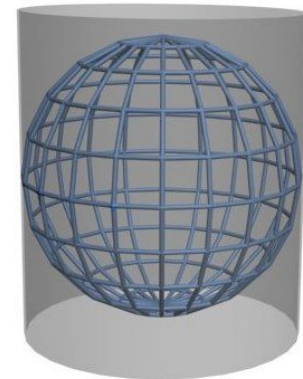
- Une carte est une projection de la terre (3D) en 2D
- Pourquoi on a besoin de projections?



(Certains) Types de projections

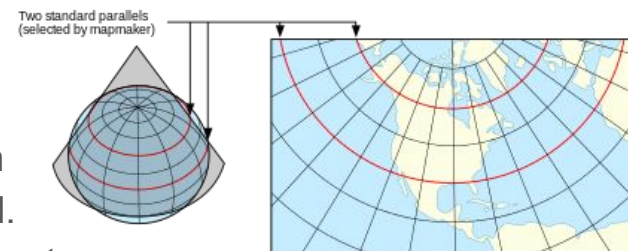
- **Cylindrique**

- le canevas ressemble à celui qui serait obtenu par projection géométrique sur un cylindre d'axe Nord-Sud.
- les méridiens sont représentés par des droites parallèles équidistantes
- les parallèles sont représentés par des droites perpendiculaires aux images des méridiens mais à distances variables.
- Ex: Projection de Mercator



- **Conique**

- Le canevas ressemble à celui qui serait obtenu par projection géométrique sur un cône de sommet situé sur l'axe Nord-Sud.
- Les méridiens sont représentés par des demi-droites concourantes formant entre elles un angle constant;
- Les parallèles sont représentés par des arcs de cercles concentriques
- Ex: Projection conique de Lambert



Propriétés des projections

- Projection **équivalente** : conserve localement les aires. Utilisé surtout dans les atlas
- Projection **conforme** : conserve localement les angles, donc les formes. Utilisée surtout pour cartes marines, militaires

- Une projection ne peut pas être **à la fois conforme et équivalente**.

Datum géodésique

- Un système de coordonnées dépend d'un **datum géodésique**
- Un datum est donné par:
 - ellipsoïde (la terre n'est pas une sphère parfaite mais elle est approximée par un ellipsoïde, plusieurs choix possibles)
 - point fondamental (croisement entre ellipsoïde et la surface réelle de la terre)
 - azimut (direction du Nord)
 - méridien d'origine
- Datums standardisés:
 - **NTF (Nouvelle Triangulation de la France)**, point fondamental au Panthéon de Paris, projection de Lambert. Cartes IGN
 - **RGF (Réseau Géodésique Français)**, projection Lambert93, CC42-50. Successeur de NTF à partir de 2001.
 - **ED50 (European Datum)**, point fondamental en Allemagne à Potsdam, projection de Mercator (UTM).
 - **WGS84 (World Geodetic System)**, système américain, projection UTM.

Les fichiers de données géographiques

- Il existe plusieurs format, ici nous allons voir certains format qui ont connu une diffusion assez importante:
 - **Shapefile**
 - **GeoJSON**
 - **KML**

Shapefile

- extension .shp
 - Accompagné par des fichiers auxiliaires (.dbf, .shx) qui contiennent informations complémentaires aux données géo et l'indexation pour un accès efficace
- format de fichier pour les systèmes d'informations géographiques (**SIG**)
- Il contient toute l'information liée à la géométrie des objets décrits, qui peuvent être :
 - des points ;
 - des lignes ;
 - des polygones.
- format binaire (pas texte)

Shapefile en R

- Lecture d'un shapefile en RStudio
- Avec la lib **sf** <https://r-spatial.github.io/sf/>
 - (autres lib utiles: cartography, cartogram, readxl)

```
Reading layer `admin-departement' from data source `/Users/dbuscaldi/Teaching/STID/departement-93/admin-departement.shp' using driver `ESRI Shapefile'
Simple feature collection with 1 feature and 4 fields
geometry type: POLYGON
dimension: XY
bbox: xmin: 2.288336 ymin: 48.80723 xmax: 2.6033 ymax: 49.01233
epsg (SRID): 4326
proj4string: +proj=longlat +datum=WGS84 +no_defs
```

GeoJSON

- Norme JSON
- Il permet de décrire des données de type point, ligne, chaîne de caractères, polygone, ainsi que des ensembles et sous-ensembles de ces types de données et d'y ajouter des attributs d'information qui ne sont pas spatiales.
- Accès aux fichiers avec la lib **geojsonR**:

```
library(geojsonR)
file_js = FROM_GeoJson(url_file_string = "feature_collection.geojson")
```

GeoJSON

Exemple de fichier GeoJSON:

```
{
  "type": "FeatureCollection",
  "features": [{
    "type": "Feature",
    "geometry": {
      "type": "Point",
      "coordinates":
        [102.0, 0.5]
    },
    "properties": {
      "prop0": "value0"
    }
  }, {
    "type": "Feature",
    "geometry": {
      "type": "LineString",
      "coordinates": [
        [102.0, 0.0],
        [103.0, 1.0],
        [104.0, 0.0],
        [102.0, 0.0]
      ]
    },
    "properties": {
      "prop0": "value0",
      "prop1": 0.0
    }
  }
  ]
}
```

KML

Lire les .kml en R c'est parfois difficile à cause des codages

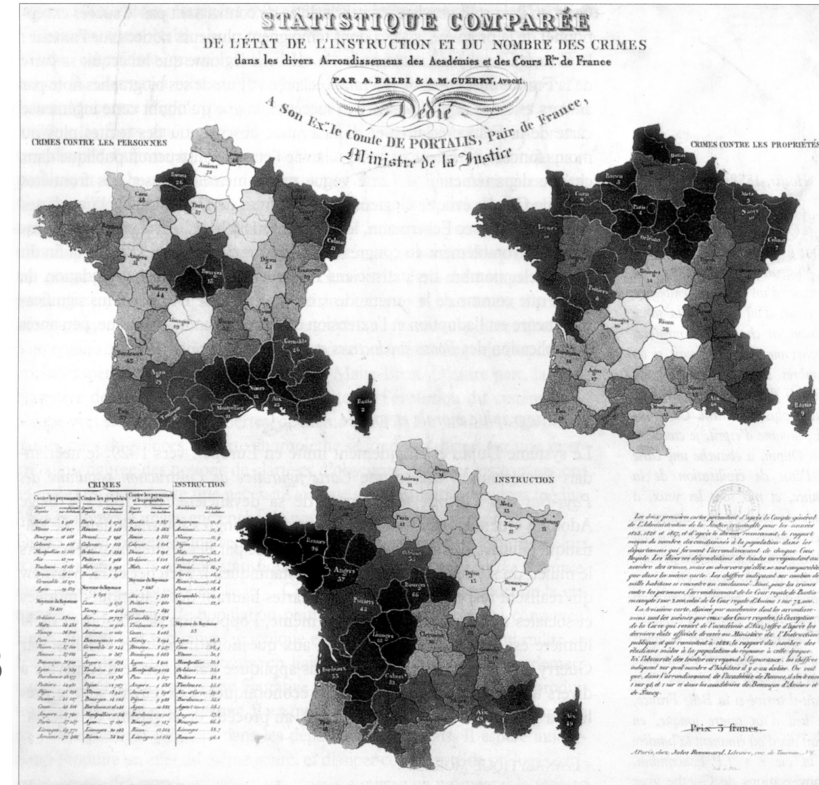
la lib est **rgdal**
avec la fonction **readOGR**

- Keyhole Markup Language
- Standard défini par OGC
- Basé sur XML
- Utilisé dans Google Maps, Google Earth
- Exemple:

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Document>
    <Placemark>
      <name>New York City</name>
      <description>New York City</description>
      <Point>
        <coordinates>-74.006393,40.714172,0</coordinates>
      </Point>
    </Placemark>
  </Document>
</kml>
```

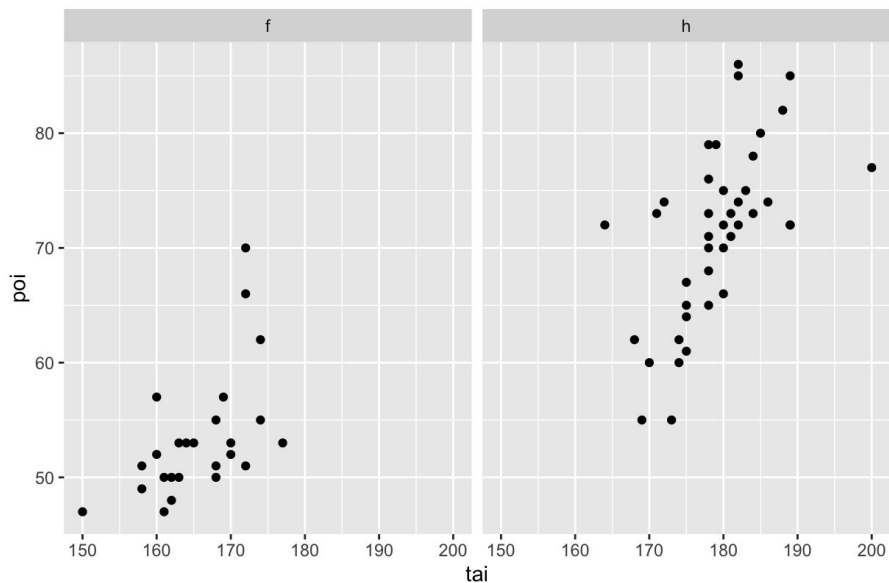
Les cartes choroplèthes

- Une **carte choroplèthe** est une carte thématique où les régions sont colorées ou remplies d'un motif qui montre une mesure statistique
- Ce type de carte facilite la comparaison d'une mesure statistique d'une région à l'autre ou montre la variabilité de celle-ci pour une région donnée
- Son inventeur est le Baron Charles Dupin, qui l'avait nommée « Carte teintée » en 1828

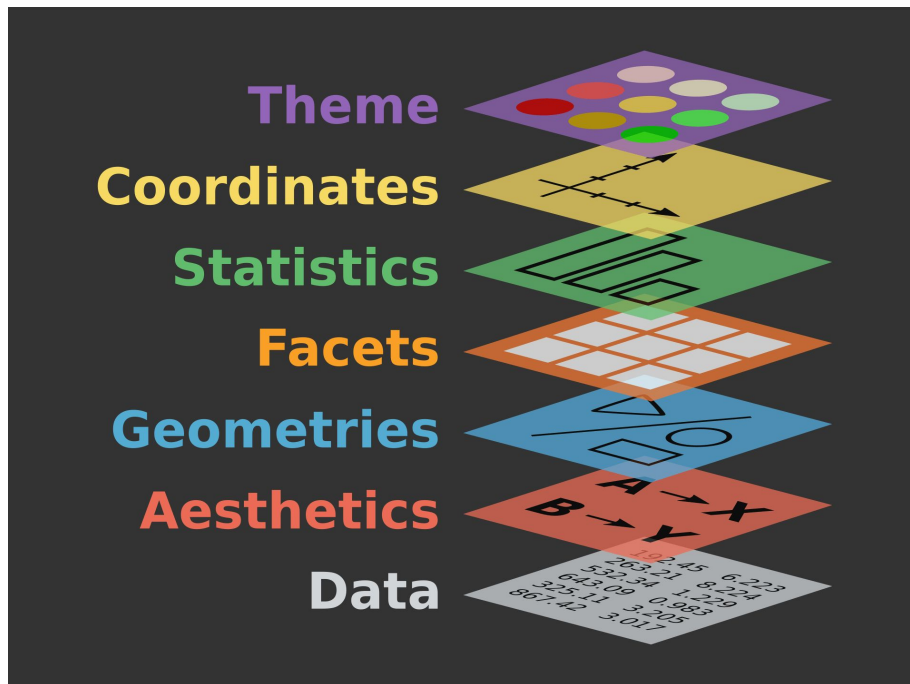


Cartes choroplèthes en R

- Plusieurs façon de générer des cartes
- Nous allons étudier en particulier la visualisation avec la lib **ggplot**
- ggplot permet de réaliser des graphiques plus complexes que avec plot, avec une syntaxe plus simple
- Exemple:
 - `qplot(tai, poi, data = t3var) +`
 - `+ facet_wrap(~ sexe)`



Les couches ggplot



Ggplot

- L'exemple antérieur montre l'utilisation de **qplot** (“quick plot”), une version simplifiée de ggplot
- Ggplot utilise une notion de “couches” de travail
- Pour voir le fonctionnement en détail, on va refaire le même avec ggplot
- D'abord on crée une base de travail comme ici:

```
p <- ggplot(data = t3var, aes(y = poi, x = tai))
```

- Ensuite on ajoute un objet géométrique, **geom_point**, qui va projeter, sur le graphique, des points aux coordonnées précédemment définies, et divisons le graphique en projetant les points de chaque sexe dans une facette différente du graphique:

```
p + geom_point() +  
  facet_grid(. ~ sexe)
```

Aesthetics (Aes)

<https://thinkr.fr/pdf/ggplot2-french-cheatsheet.pdf>

- La fonction aes nous permet de modifier l'esthétique des graphiques
- Exemple: couleurs conditionnels

```
p + geom_point() +  
  facet_grid(. ~ sexe) + +  
  aes(color = poi > 60)
```

```
ggplot(iris, aes(Sepal.Length, Sepal.Width)) +  
  + geom_point(aes(color = Species))
```

- Autres options esthétiques:
 - shape (forme) : aes(shape = ...)
 - fill (remplissage) : aes(fill = ...)
 - linetype (type de ligne) : aes(linetype = ...)
 - label (étiquettes) : aes(label = ... , vjust=...)