

Correction TP1

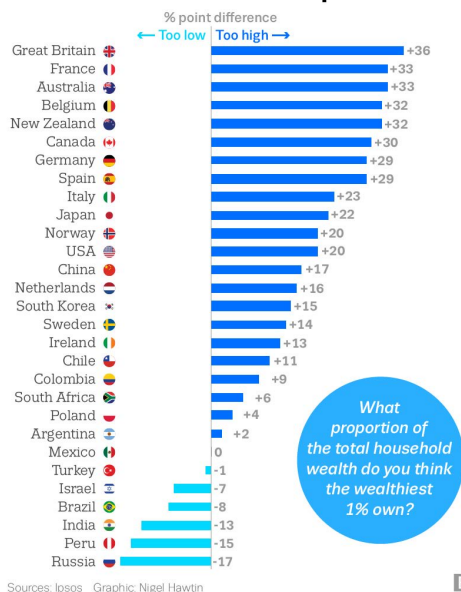
Arthur Katosky

18/01/2019

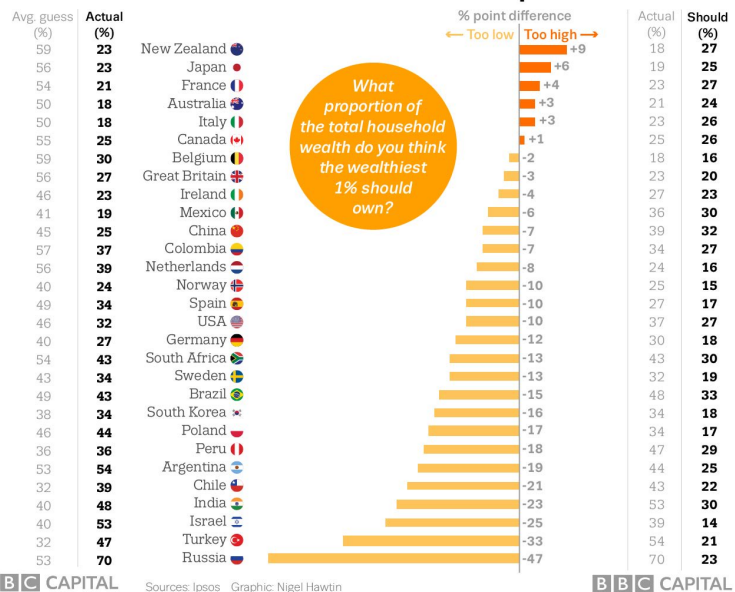
1. Rappel de la consigne

1. Produisez une critique constructive du graphique suivant
2. Proposez une visualisation alternative des données (data/how-much-the-rich.csv)

How much do the super rich own?



How much should the super rich own?



Source.

2. Proposition de correction

2.1 Critique constructive

Rappel: Nous essayons de répondre aux 3 questions suivantes:

- la question à laquelle répond le graphe est-elle claire? (Q)
- les données mobilisées permettent-elles de répondre à la question posée? (D)
- la représentation visuelle choisie utilise-t-elle correctement les données pour répondre à la question? (V)

Q: Les deux questions posées (How much *do* the rich own? How much *should* the rich own?) sont claires et intéressantes. Le fait que le titre reste une question (et non pas une affirmation) doit nous alerter! Y a-t-il seulement un message à communiquer?

D: Les données représentées ne permettent pas de répondre à la question posée. Les données répondent à deux questions, la seconde est liée — mais pas explicitement — aux questions initiales: quelle est la différence entre ce que nous *souhaiterions* que les riches possèdent et ce qu'ils possèdent réellement? La question du premier graphique, en revanche, est complètement hors sujet: quelle est la différence entre ce que nous *supposons* que les riches possèdent et ce qu'ils possèdent réellement?

En résumé, le travail sur les données (calculer une différence) ne permet pas de répondre aux questions initiales. Il est ainsi impossible de comparer graphiquement “how much the rich own” ou “how much the rich should own” dans différents pays.

V:

Positif:

- Utilisation d'un graphique en barre (très efficace pour véhiculer de l'information quantitative).
- La couleur renforce l'opposition positif vs. négatif. (très efficace pour véhiculer une information qualitative)
- Utilisation des drapeaux pour faciliter le repérage des pays, qui ne sont pas classés dans l'ordre alphabétique.
- Tentative de guide pour la lecture de la légende.

Négatif:

- Les mêmes données sont répétées deux fois (“Actual”) sur les deux graphes.
- Il y a beaucoup trop de nombres sur les graphiques. Le graphique passe le test d'autosuffisance pour les nombres sur les barres, mais les nombres sur la droite n'ont pas de contre-partie graphique. Et ils sont très difficilement lisibles puisque non ordonnés.

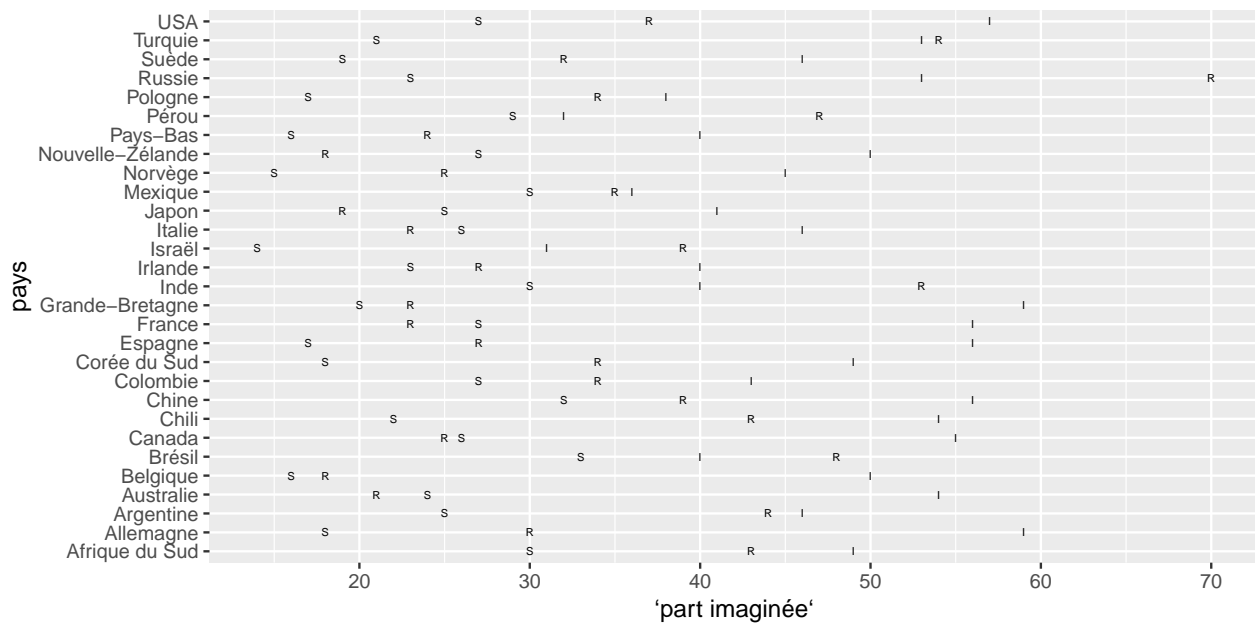
Autres remarques:

- “Super rich” est un peu exagéré pour parler des 1%. (Pour un ordre de grandeur, en France, les 1% les plus riches possèdent un capital de env. 2 000 000 €.)
- À cause du décalage entre la question posée et les données représentées, le designer se sent obligé de préciser la question exacte qui a été posée dans un cartouche de couleur ; *a priori* on s'attend à ce que ces cartouches très visibles contiennent un élément narratif fondamental pour comprendre / interpréter le graphique.

2.2 Proposition A

Il est en réalité possible de représenter toutes les données dans le même graphique, voir sur le même axe, puisque toutes les données sont homogènes. Voyons ce que cela donne:

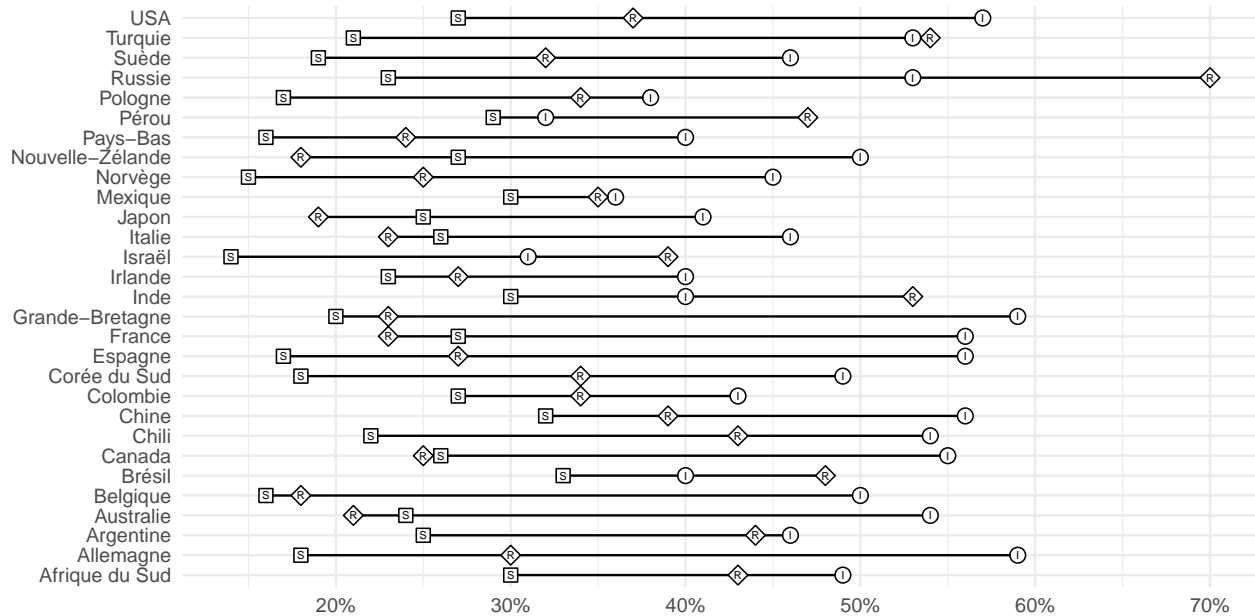
```
data <- read_csv("data/how-much-the-rich.csv")
data %>%
  ggplot(aes(y=pays)) +
  geom_point(aes(x=`part imaginée`), shape='I') +
  geom_point(aes(x=`part souhaitée`), shape='S') +
  geom_point(aes(x=`part réelle`), shape='R')
```



On peut améliorer rapidement en:

- enlevant les titres des axes
- mettant les pourcentages sur l'axe des x
- reliant les points entre eux pour souligner les longueurs
- ajoutant des formes pour souligner les différences de lettre (redondance)

```
data %>%
  mutate(
    pct_min = pmin(`part imaginée`, `part souhaitée`, `part réelle`),
    pct_max = pmax(`part imaginée`, `part souhaitée`, `part réelle`)
  ) %>%
  ggplot(aes(y=pays)) +
  geom_segment(aes(x=pct_min, xend=pct_max, yend=pays)) +
  geom_point(aes(x=`part imaginée`, shape=21, size=3, fill="white")) +
  geom_point(aes(x=`part imaginée`, shape='I')) +
  geom_point(aes(x=`part souhaitée`, shape=22, size=3, fill='white')) +
  geom_point(aes(x=`part souhaitée`, shape='S')) +
  geom_point(aes(x=`part réelle`, shape=23, size=3, fill='white')) +
  geom_point(aes(x=`part réelle`, shape='R')) +
  scale_x_continuous(labels=scales::number_format(suffix='%')) +
  theme_minimal() +
  labs(
    x = NULL,
    y = NULL
  )
```



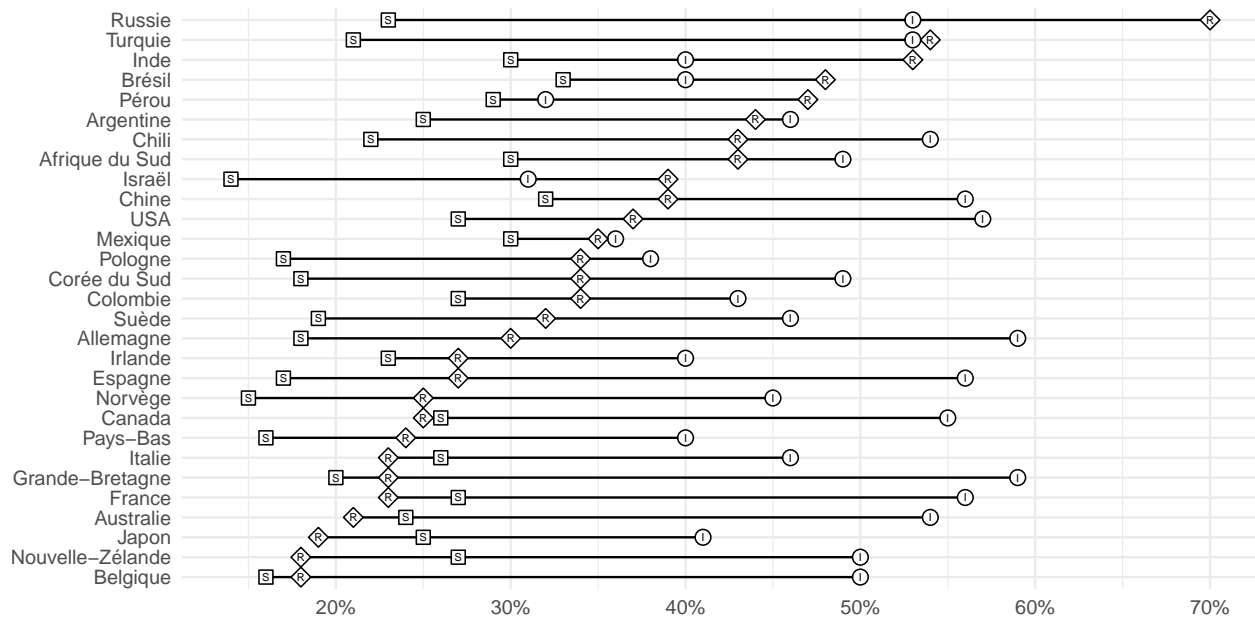
Nous pouvons être content.

Reste la question de la scénarisation. Aucun message ne se dégage de ce graphique! La visualisation initiale essayait de mettre en lumière:

1. l'opposition entre la réalité et la perception et
2. l'opposition entre la réalité et la réalité souhaitée.

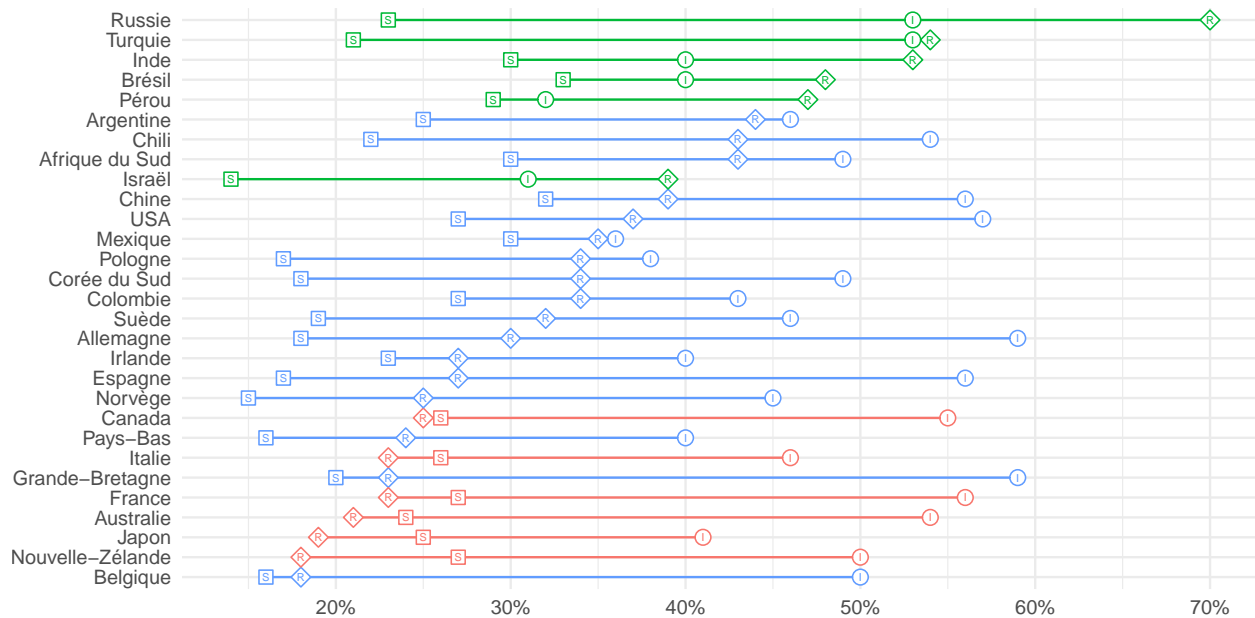
Le facteur commun aux deux approches est donc la distribution réelle de la richesse. Et si nous triions nos barres du plus petit au plus grand R? (du pays le plus égalitaire au plus inégalitaire)

```
data %>%
  mutate(
    pct_min = pmin(`part imaginée`, `part souhaitée`, `part réelle`),
    pct_max = pmax(`part imaginée`, `part souhaitée`, `part réelle`),
    pays = fct_reorder(pays, `part réelle`)
  ) %>%
  ggplot(aes(y=pays)) +
  geom_segment(aes(x=pct_min, xend=pct_max, yend=pays)) +
  geom_point(aes(x=`part imaginée`, shape=21, size=3, fill="white")) +
  geom_point(aes(x=`part imaginée`, shape='I')) +
  geom_point(aes(x=`part souhaitée`, shape=22, size=3, fill='white')) +
  geom_point(aes(x=`part souhaitée`, shape='S')) +
  geom_point(aes(x=`part réelle`, shape=23, size=3, fill='white')) +
  geom_point(aes(x=`part réelle`, shape='R')) +
  scale_x_continuous(labels=scales::number_format(suffix='%')) +
  theme_minimal() +
  labs(
    x = NULL,
    y = NULL
  )
```



En revanche, le graphique original souligne clairement l'opposition entre les perception excessives ou non (bleu vs. bleu clairs) et les morales correctives dans un sens ou dans l'autre (organge vs. orange clair). L'ordre des lettres I, S et R est très évocateur, et on peut facilement utiliser la couleur pour souligner les 3 catégories présentes dans les données.

```
data %>%
  mutate(
    pct_min = pmin(`part imaginée`, `part souhaitée`, `part réelle`),
    pct_max = pmax(`part imaginée`, `part souhaitée`, `part réelle`),
    pays = fct_reorder(pays, `part réelle`)
  ) %>%
  ggplot(aes(y=pays, color=case_when(
    `part imaginée` < `part réelle` ~ "S < I < R",
    `part réelle` < `part souhaitée` ~ "R < S < I",
    TRUE ~ "S < R < I"
  ))) +
  geom_segment(aes(x=pct_min, xend=pct_max, yend=pays)) +
  geom_point(aes(x=`part imaginée`, shape=21, size=3, fill="white")) +
  geom_point(aes(x=`part imaginée`, shape='I')) +
  geom_point(aes(x=`part souhaitée`, shape=22, size=3, fill='white')) +
  geom_point(aes(x=`part souhaitée`, shape='S')) +
  geom_point(aes(x=`part réelle`, shape=23, size=3, fill='white')) +
  geom_point(aes(x=`part réelle`, shape='R')) +
  scale_x_continuous(labels=scales::number_format(suffix='%')) +
  theme_minimal() +
  guides(color="none") +
  labs(
    x = NULL,
    y = NULL
  )
```

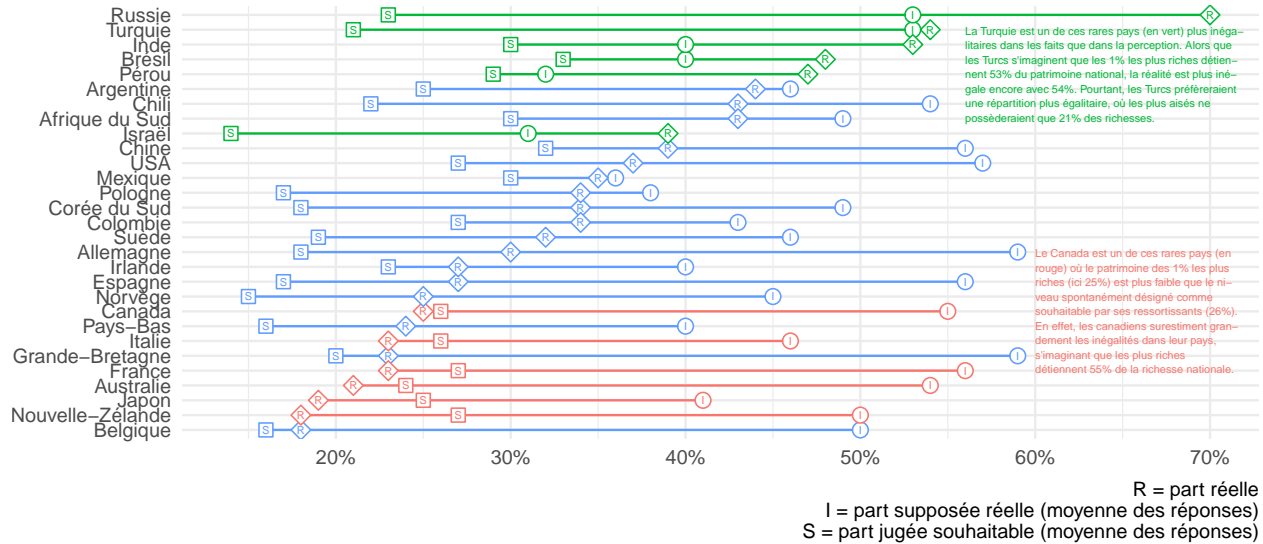


Enfin, il nous reste à donner un titre, et ajouter des aides de lecture (parce que le graphique n'est pas simple à lire).

```
data %>%
  mutate(
    pct_min = pmin(`part imaginée`, `part souhaitée`, `part réelle`),
    pct_max = pmax(`part imaginée`, `part souhaitée`, `part réelle`),
    pays = fct_reorder(pays, `part réelle`)
  ) %>%
  ggplot(aes(y=pays, color=case_when(
    `part imaginée` < `part réelle` ~ "S < I < R",
    `part réelle` < `part souhaitée` ~ "R < S < I",
    TRUE ~ "S < R < I"
  ))) +
  geom_segment(aes(x=pct_min, xend=pct_max, yend=pays)) +
  geom_point(aes(x=`part imaginée`, shape=21, size=3, fill="white")) +
  geom_point(aes(x=`part imaginée`, shape='I')) +
  geom_point(aes(x=`part souhaitée`, shape=22, size=3, fill='white')) +
  geom_point(aes(x=`part souhaitée`, shape='S')) +
  geom_point(aes(x=`part réelle`, shape=23, size=3, fill='white')) +
  geom_point(aes(x=`part réelle`, shape='R')) +
  scale_x_continuous(labels=scales::number_format(suffix='%')) +
  annotate("text", x=60, y='Canada', size=1.6, hjust=0, color='#F8766D', label="Le Canada est un de ces")
  annotate("text", x=56, y=25, size=1.6, hjust=0, color='#00BA38', label="La Turquie est un de ces rares")
  theme_minimal() +
  guides(color="none") +
  labs(
    x = NULL,
    y = NULL,
    title = "Des croyances indépendantes de la réalité?",
    subtitle = "Part des richesses nationales détenues par les 1% les plus riches",
    caption = "R = part réelle\nI = part supposée réelle (moyenne des réponses)\nS = part jugée souhaitable"
  )
```

Des croyances indépendantes de la réalité?

Part des richesses nationales détenues par les 1% les plus riches

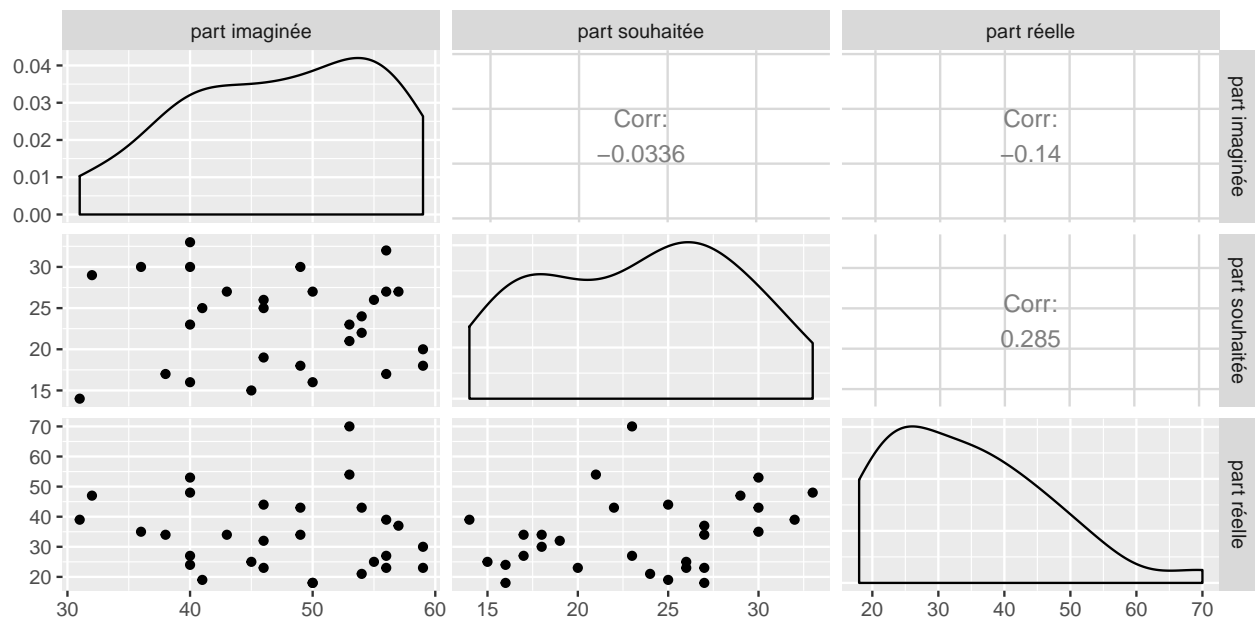


Mais le principal problème de l'analyse, c'est qu'il n'y a pas vraiment de choses à voir (d'où le titre proposé). Les taux imaginés et souhaités sont indépendants l'un de l'autre, et quasi-indépendants des inégalités réelles. Peut-être après tout est-ce que le resenti des inégalités ne se joue pas au niveau des 1%?

2.2 Proposition B

Plutôt que de présenter les 3 variables sur le même axe comme précédemment, on peut vouloir les représenter avec chaque variable sur son axe. Problème: on est en 2D (et de toute façon la 3D a de très mauvaises propriétés cognitives). On peut commencer par opposer les variables 2 à deux pour voir.

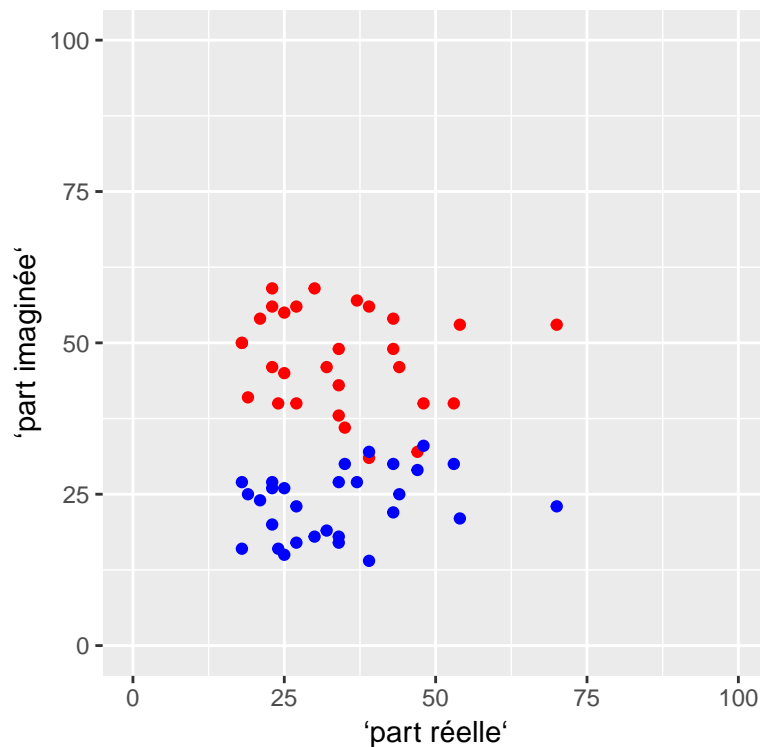
```
data %>%
  select(`part imaginée`, `part souhaitée`, `part réelle`) %>%
  ggpairs
```



Pas grand chose d'intéressant. Du point de vue graphique, l'interprétation est handicapée par les axes qui ne commencent pas tous au même endroit. Il vaudrait mieux que tous les axes commencent à 0% et se terminent à 100%. (Ainsi, en plus de la position, on a la longueur depuis les bords du graphique qui a un sens.)

Si on veut opposer une des grandeurs, le plus pertinent est d'opposer la réalité (part réel) vs. les choses imaginaires (part supposée, part souhaitée).

```
data %>%
  ggplot(aes(x=`part réelle`)) +
  geom_point(aes(y=`part imaginée`), color='red')+
  geom_point(aes(y = `part souhaitée`), color='blue')+
  scale_y_continuous(lim=c(0, 100))+
  scale_x_continuous(lim=c(0, 100))+
  coord_equal()+
  guides(color=FALSE)
```



Rajouter la diagonale permet de faire de diagnostiquer lorsque les grandeurs “imaginaires” dépassent les grandeurs réelles. On peut faire d'autres améliorations (formats, arrière plan, titre des axes, couleurs).

```
data %>%
  ggplot(aes(x=`part réelle`)) +
  geom_point(aes(y=`part imaginée`), color='red')+
  geom_point(aes(y = `part souhaitée`), color='blue')+
  geom_abline(a=1, b=0, linetype=2) +
  scale_y_continuous(lim=c(0, 100), labels=scales::number_format(suffix='%'))+
  scale_x_continuous(lim=c(0, 100), labels=scales::number_format(suffix='%'))+
  theme_minimal() +
  coord_equal()+
  guides(color=FALSE) +labs(
    x = "Part réelle",
    y = "Part supposée\nPart souhaitée",
    title = "Des croyances indépendantes de la réalité",
```



```

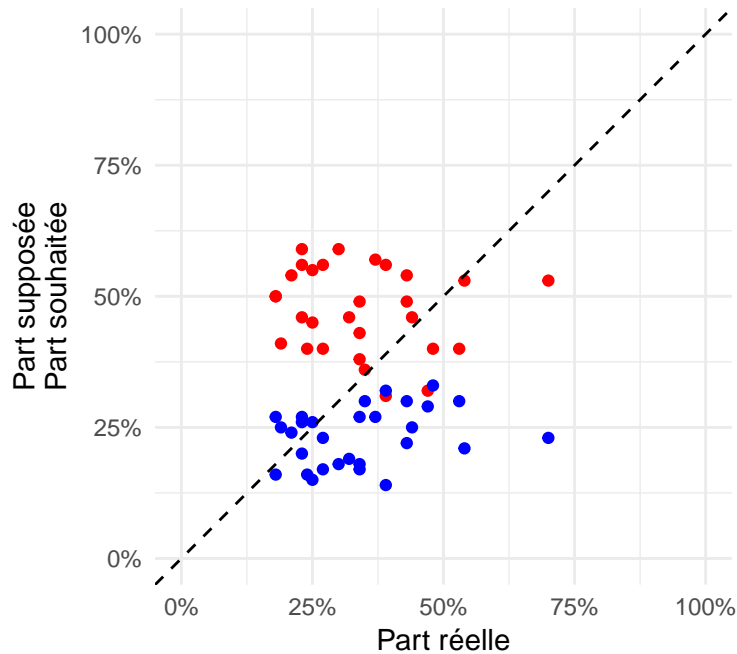
    subtitle = "Part des richesses nationales détenues par les 1% les plus riches"
)

```

Warning: Ignoring unknown parameters: a, b

Des croyances indépendantes de la réalité

Part des richesses nationales détenues par les 1% les plus riches



On voit clairement l'absence de lien entre la part réelle et ce que les gens pensent. Restent quelques problèmes. Il resterait à traiter: l'identification entre part supposée / souhaitée et une couleur (facile: par exemple en faisant suivre "Part supposée réelle" d'un marqueur) et l'étiquetage des points (difficile: il n'est pas possible de relier deux points représentant un même pays sans rendre le graphique illisible ; et il y a trop de valeurs pour pouvoir utiliser la couleur et *a fortiori* la forme pour distinguer les pays). Une aide de lecture serait bienvenue (les points rouges sous la diagonale montrent une sous-estimation de la richesse ; les points bleus au-dessus de la diagonale sont les cas où la part jugée souhaitable est supérieure à la part réelle).