

# Візуалізація даних як навичка

Андрій Газін, [Textura.in.ua](https://textura.in.ua)

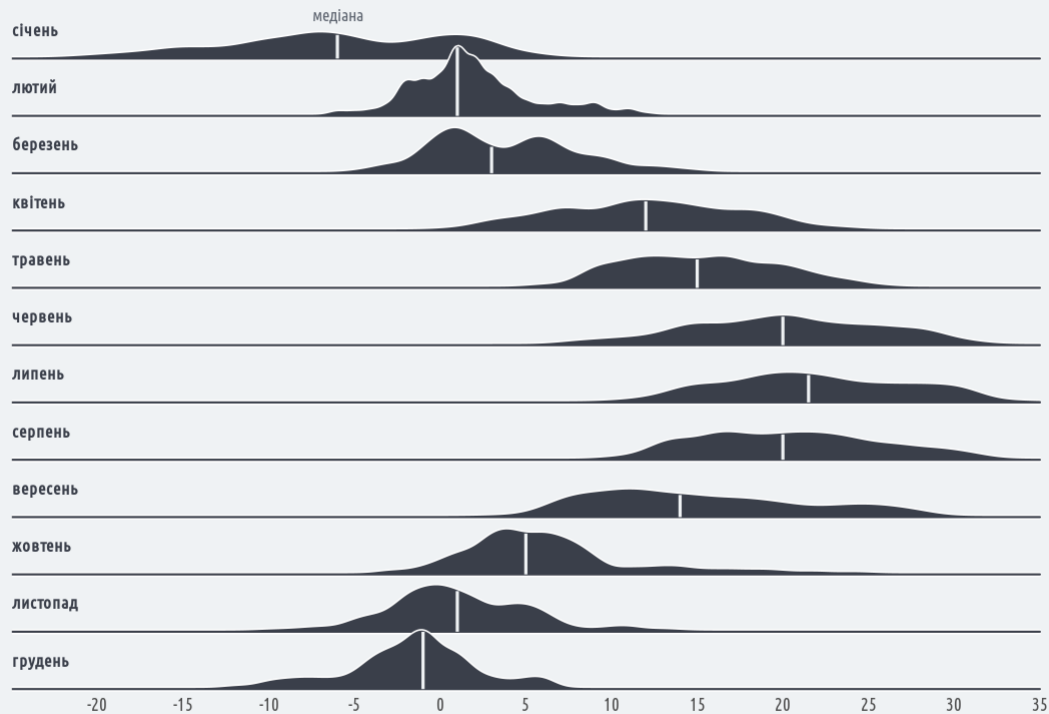
Я роблю  
красиві  
графіки



такі,  
наприклад

## Температура в Києві

Розподіл за місяцями у 2016 році

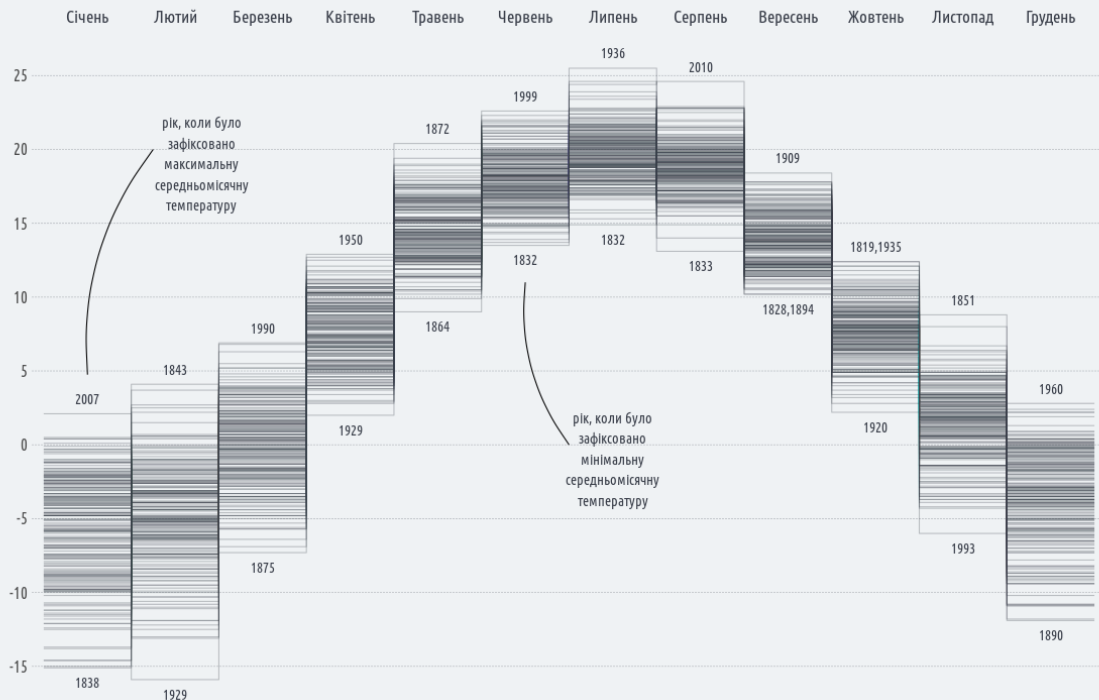


Дані: Weather Underground | Візуалізація: Textura.in.ua

або такі

## Температура в Києві, 1812-2014 рік

Горизонтальні лінії на графіку позначають середню температуру для кожного місяця за період спостережень з 1812 до 2014 року



**або такі**  
[наберіться  
терпіння]

## Частотність літер залежно від позиції у слові

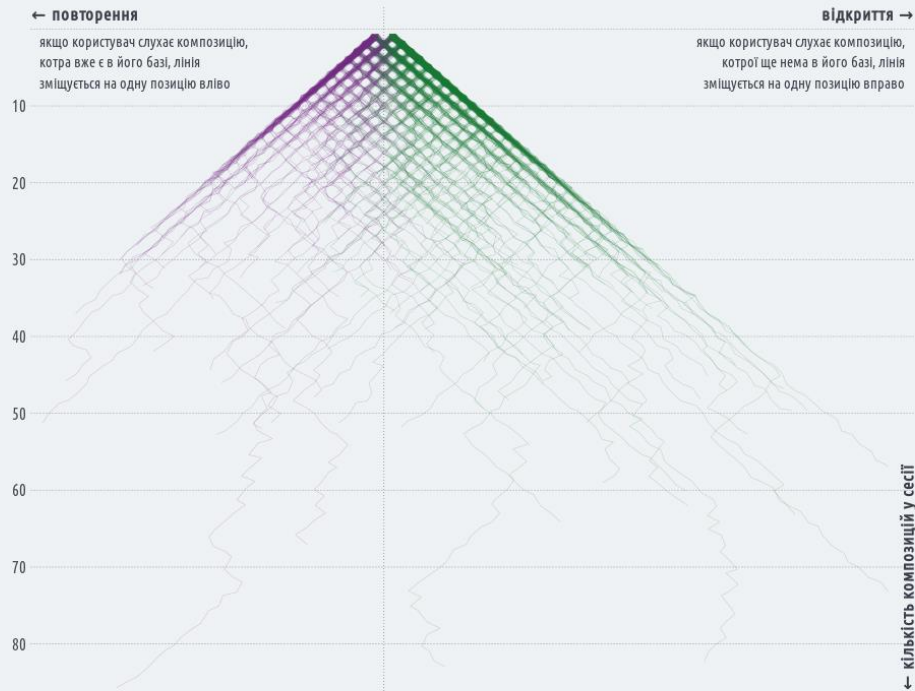
Дані нормалізовані за довжиною слова. Для обчислення використано 153.2 тисячі слів зі словника української мови



або такі  
[в мене їх  
багато]

## Статистика прослуховування музики

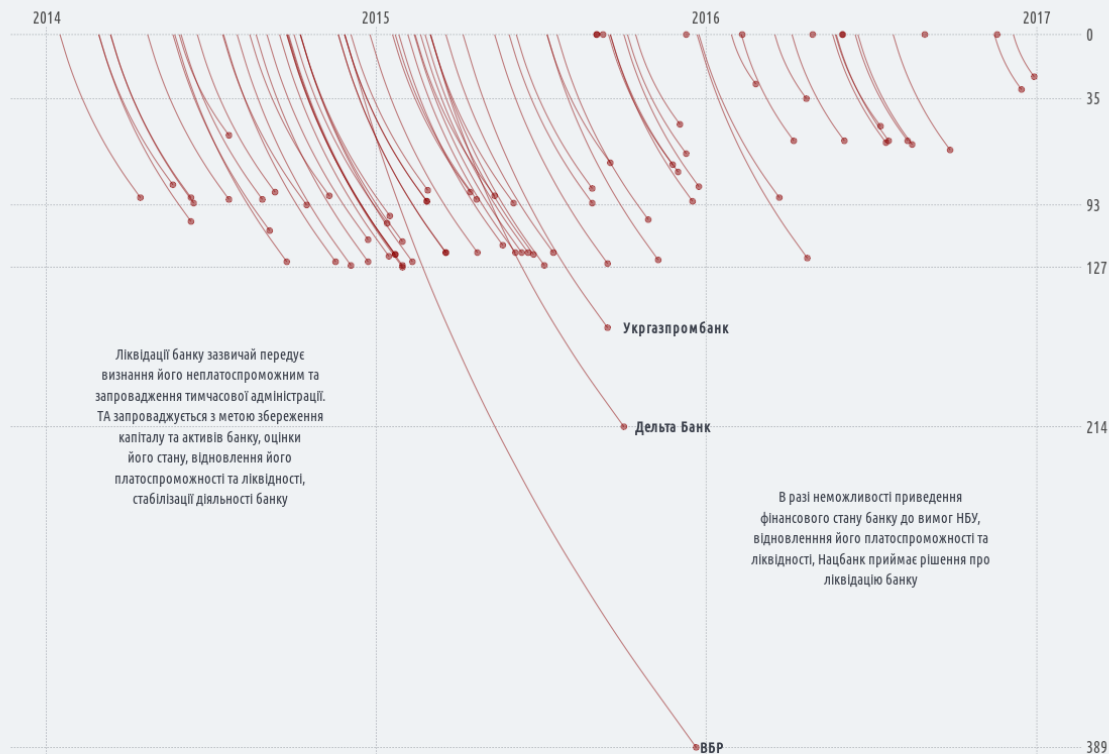
Кожна лінія на графіку - одна сесія прослуховування музики. Довжина лінії позначає кількість композицій у сесії. Зелений колір лінії означає, що в сесії переважають нові композиції. Фіолетовий колір лінії означає, що в сесії переважають вже відомі користувачу композиції



# або такі [ага, ще один]

## Хроніка чистки банківської системи України

Початок лінії позначає дату визнання банку неплатоспроможним. Точка позначає дату прийняття рішення про ліквідацію банку. Довжина лінії позначає кількість днів між визнанням банку неплатоспроможним та прийняттям рішення про його ліквідацію

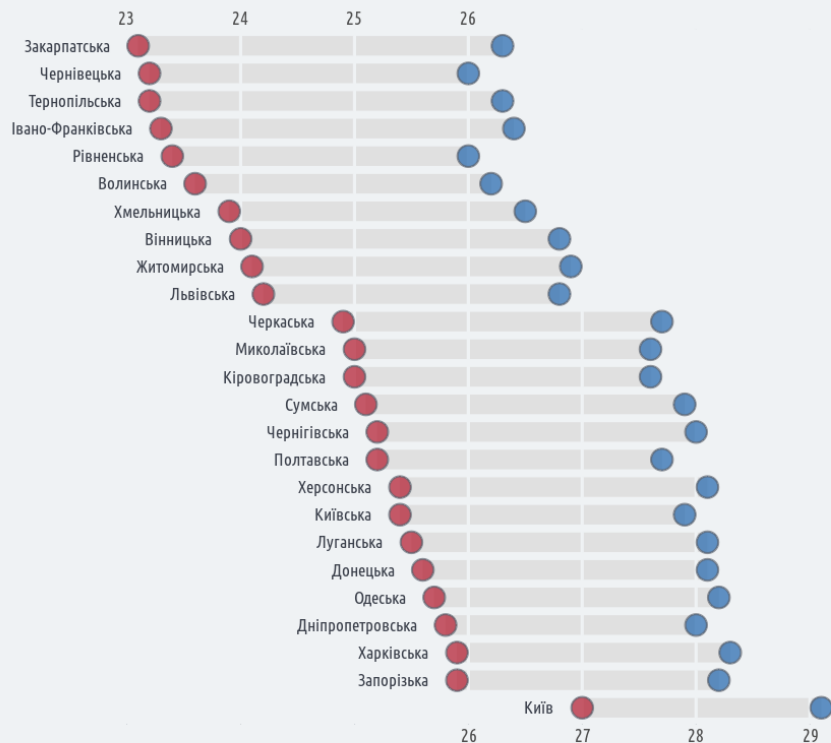


Дані: НБУ | Візуалізація: textura.in.ua

або такі  
[це останній,  
чесне слово]

## Середній вік першого шлюбу

В розрізі регіонів та статей. ● жінки ● чоловіки



Дані: Державна служба статистики, 2015 рік | Візуалізація: Textura.in.ua



**Ці графіки  
базуються на  
певних  
принципах.  
Зокрема...**

- Досліджувати дані за допомогою візуалізації
- Шукати нові форми візуалізації даних
- Експериментувати із елементами візуалізації (сіткою, вісями, підписами)

# Здається, ці принципи працюють

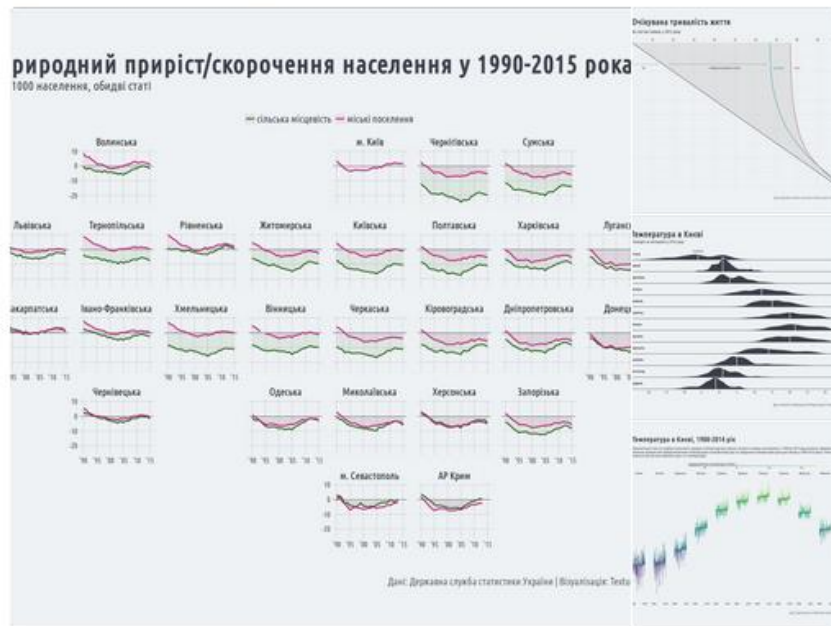
Hadley Wickham ретвітнув(ла)



**Duc Quang Nguyen** @duc\_qn · 26 черв.

I don't speak Ukrainian, but these #rstats charts are gorgeous [buff.ly/2tMeyrd](https://buff.ly/2tMeyrd)  
Thanks @AndriyGazin to provide the R code 🙌 #dataviz

Перекласти (англійська)



# Я також викладаю візуалізацію даних

- Магістерська програма з Data Science в УКУ
- UCU Data Science Summer School
- Тренінги, воркшопи, мітапи

**Кожного разу  
я маю дати  
собі відповідь  
на кілька  
питань**

- ~~- В чому сенс життя?~~
- ~~- Хто ми і навіщо ми тут?~~
- ~~- Чи ми самотні у цьому всесвіті?~~
- Які знання та навички я хочу передати?

# Розповідати про різні інструменти - not much fun [але куди без цього]

Вибір мови, бібліотеки,  
фреймворку чи  
програми залежить від  
багатьох факторів (наявні  
дані, поставлені задачі,  
workflow і т.д.)

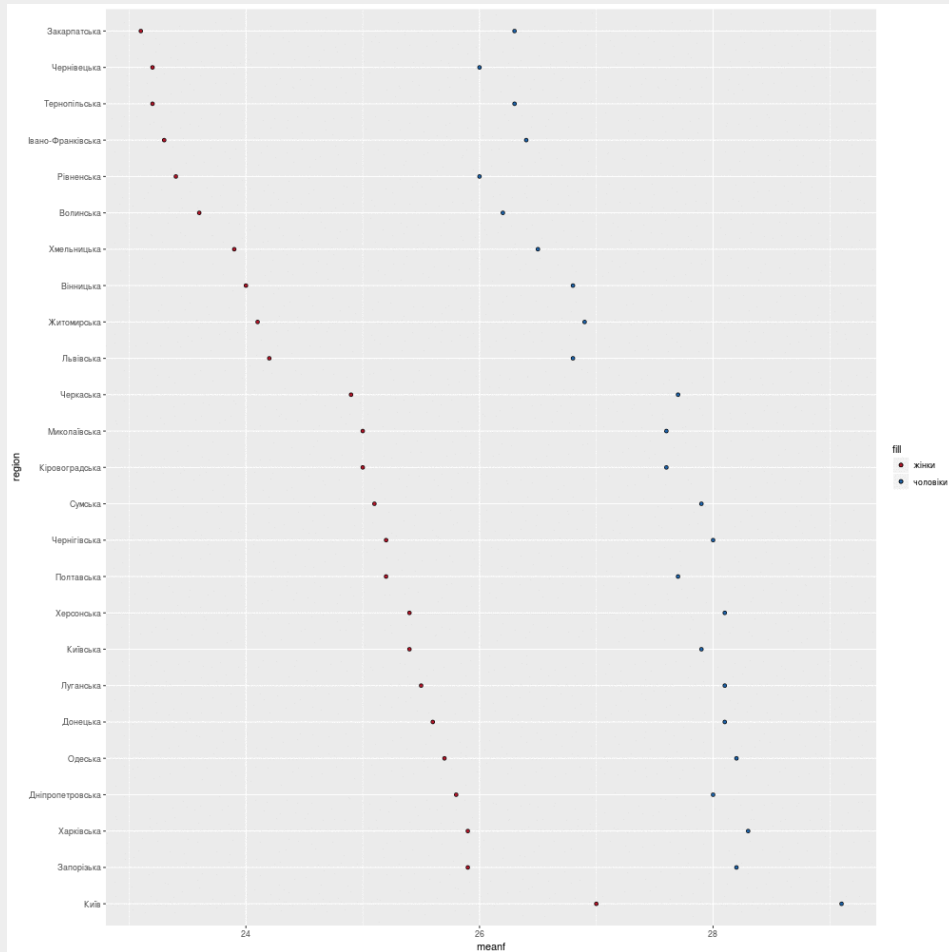
# Розповідати про типи графіків та їх застосування - трохи нудно [хоча й треба]

Візуалізація даних це не обов'язково про типи графіків. Скоріше про пошук найефективнішого способу візуального кодування інформації

**Пояснювати,  
як писати код,  
- не надто  
цікаво  
[але, звісно, потрібно]**

QKHKFHKNJ**R**HKNFMGKR  
PKJGHKTPOLS,LPPYJEPVX  
,NFMBJWHGUEGWIPQOP  
**T**OHIYRPTYODFDGJDLTI  
TJFHDKGHDKFJGHTIURIO  
RJILGLK**F**JGLHPOWUEQJY  
WGJHFLGGHIYOIPYLGJHL  
FGH;LK;K;ETRRU**M**JRHGTK

# Значно цікавіше говорити про дизайн- процес





**Ок, отже,  
що таке  
візуалізація  
даних?**

# З точки зору процесу

Кодування даних за допомогою візуальних засобів (позиції, довжини, площі, кольору і т.д.)

# З точки зору результату

Графічна репрезентація  
патернів і трендів у даних

# З точки зору задачі

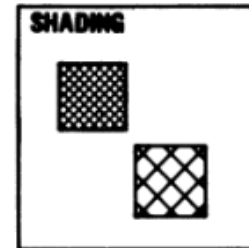
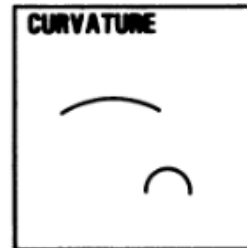
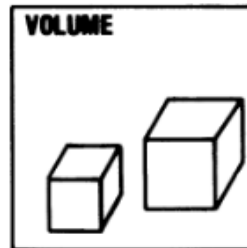
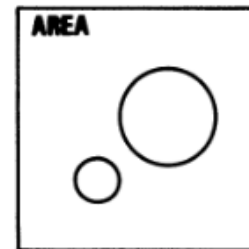
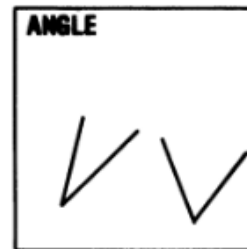
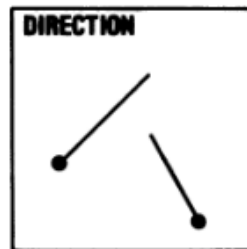
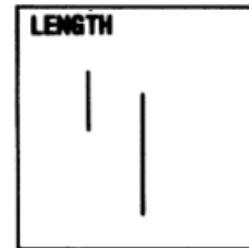
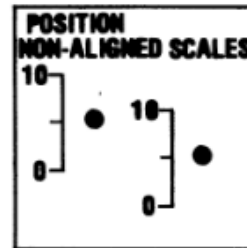
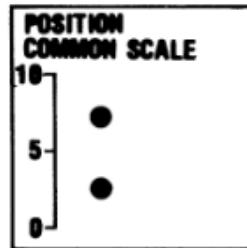
Скорочення часу та  
зусиль читача та/або  
користувача на  
сприйняття інформації

**Ок, до чого це  
все?**

# Ми маємо добре знати наші дані

Так званий провенанс даних, тобто історію їх створення та трансформації

# Ми маємо розуміти способи кодування даних



**Ми маємо  
знати  
граматику та  
семіологію  
графіки**

Вони дають нам набір елементів, з яких складається візуалізація даних, а також певні правила їх комбінування



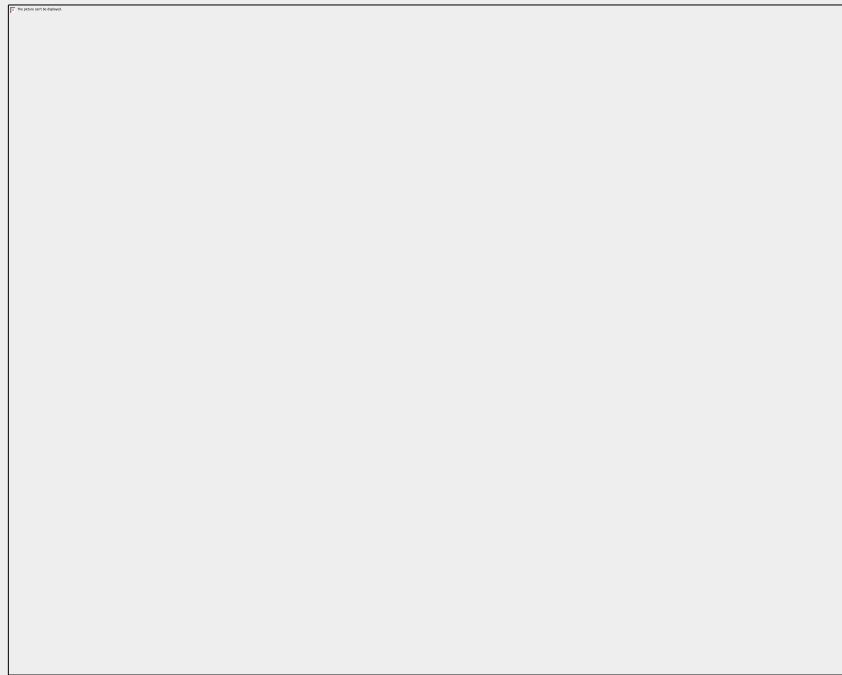
**Ми маємо  
розуміти, що  
саме ми  
хочемо  
показати за  
допомоги  
візуалізації**

Значна частина проблем із візуалізацією даних походить від того, що автор не знає, навіщо ця візуалізація потрібна, яке її призначення

# Ми маємо знайти найкращий спосіб кодування для наших даних

І це не вибір між типами графіків, це вибір між візуальними властивостями, які найкраще дозволяють проявити дані

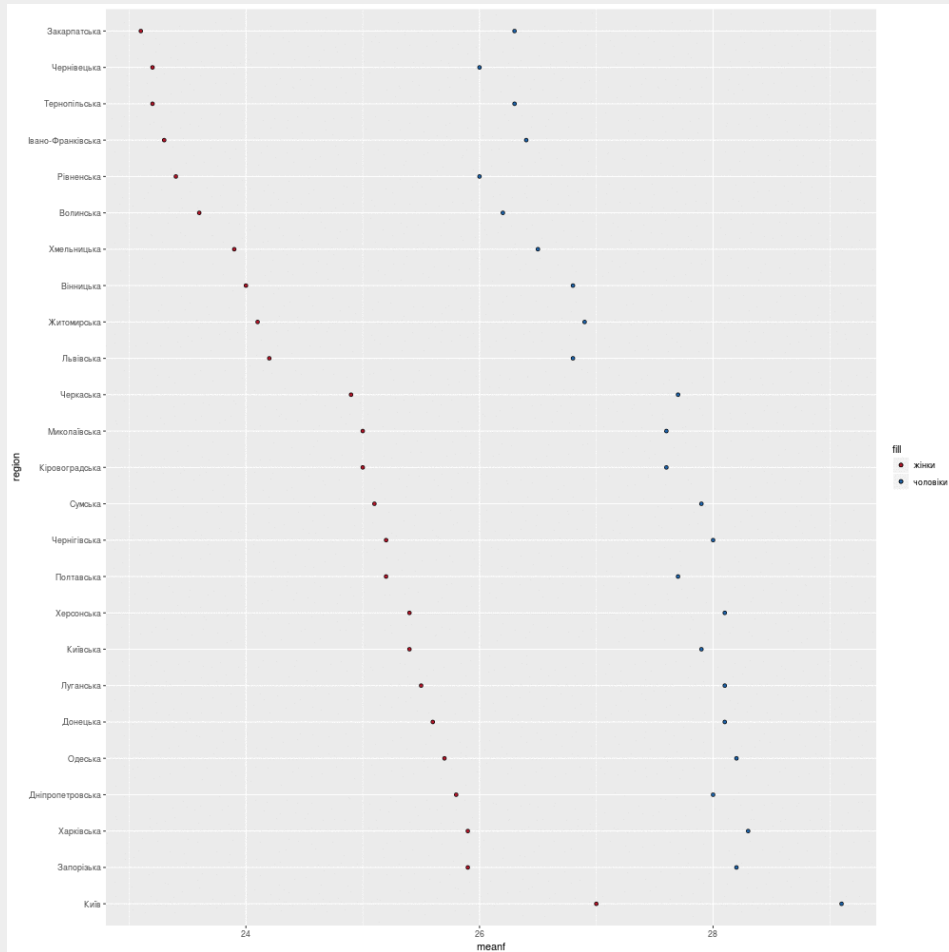
**Ми маємо  
розуміти, в  
який спосіб  
користувачі  
декодують  
візуалізацію**



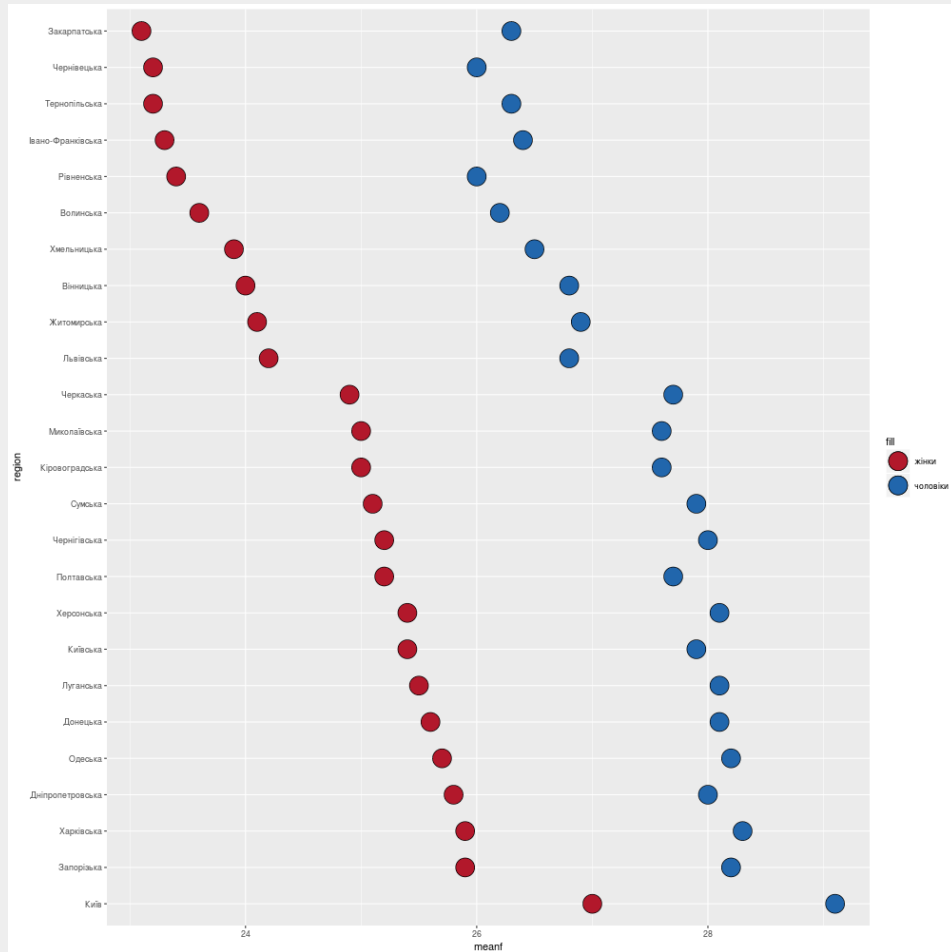
**Ми маємо  
уявляти,  
за яких умов  
відбудеться  
контакт з  
візуалізацією**

Це накладатиме певні обмеження на вибір елементів, якими ми можемо скористатись, а також власне кодування інформації

Ми маємо  
скоротити  
час і зусилля,  
необхідні на  
сприйняття  
інформації



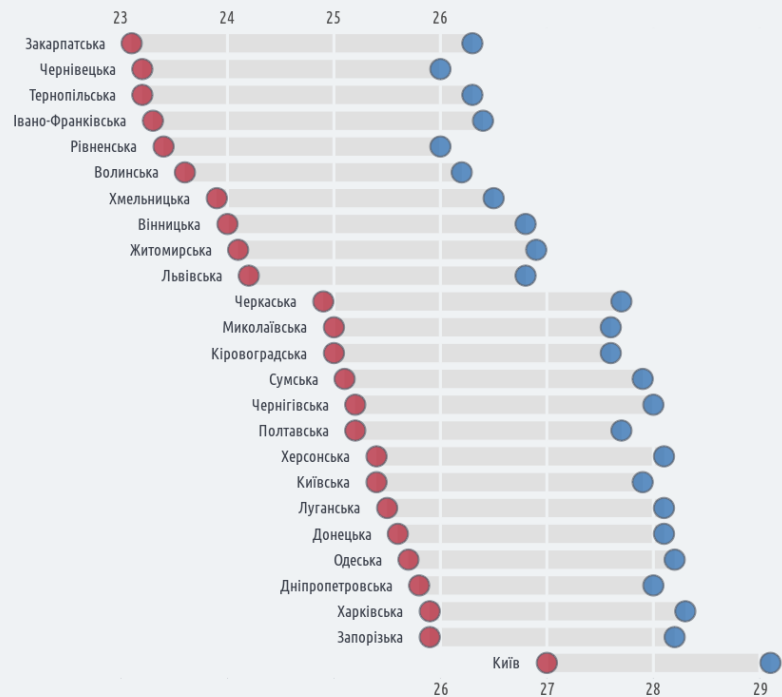
Ми маємо  
пройти шлях  
від



До

## Середній вік першого шлюбу

В розрізі регіонів та статей. ● жінки ● чоловіки



Дані: Державна служба статистики, 2015 рік | Візуалізація: Textura.in.ua

**Дякую за  
увагу!**  
Питання?







# Андрій Газін

[textura.in.ua](http://textura.in.ua)

[fb.com/textura.in.ua](https://fb.com/textura.in.ua)

[twitter.com/andriygazin](https://twitter.com/andriygazin)