

Конспект лекцій. Тиждень 1.

Мова візуалізації

Візуалізація - це представлення інформації, даних, фактів у візуальній формі. Водночас, візуалізація є мовою, в якій використовуються геометричні об'єкти - точка, лінія, частина поверхні, а також візуальні канали - колір, довжина, орієнтація, розмір. Фактично, мова візуалізації - це продовження звичайної мови, тому що тексти - її частина.

Одночасно, як і будь яка мова, її базові елементи можна комбінувати багатьма способами. Проте, не всі комбінації мають сенс. До того ж, різні типи даних вимагають різних способів їх представлення мовою візуалізації - для них потрібно використовувати різні способи візуального кодування.

З одного боку, може здатися, що це ускладнює задачу інформаційного дизайнера. Насправді, якщо знати мову візуалізації та правила, у який спосіб краще представляти ті чи інші дані, це сильно полегшує роботу - тому що обмежує кількість можливих варіантів.

Отже, саме цим ми зараз і займемося: подивимося, які типи даних існують, і як їх кодувати за допомогою цієї мови у найбільш ефективний спосіб.

Типи даних

Ми використаємо найпростішу схему класифікації. За нею, дані поділяються на три типи:

- кількісні (quantitative) - все, що можна порахувати та записати у числовій формі;
- впорядковані (ordered) - якісні дані, те, що можна розташувати у якомусь порядку - дні тижня, градації шкали оцінювання (наприклад, від "дуже погано" до "дуже добре");
- категорійні (categorical) - невпорядковані якісні дані. Практично все, що не відноситься до перших двох типів - назви країн, назви з будь яких наборів, різноманітні типи, тощо.

Елементи мови. Мітки та канали

Елементами мови візуалізації є мітки та візуальні канали. **Мітки** - це базові графічні елементи (найпростіші геометричні об'єкти):

- точка
- лінія
- площина (на 2D поверхні)
- об'ємне тіло (в 3D)

Канали - це спосіб, у який ми можемо показати наші позначки. Тобто, ми можемо контролювати як буде виглядати позначка, за допомогою таких візуальних каналів, як:

- позиція
- розмір
- форма
- орієнтація
- відтінок, насиченість, яскравість (кольору).

Послідовність дій

Отже, для візуалізації ваших даних, перше, що необхідно зробити

- це порахувати кількість змінних (наприклад, скільки колонок є у вашій таблиці з даними);
- визначити для кожної із цих змінних, до якого типу даних вона відноситься: до кількісних, впорядкованих чи категорійних.

Після цього, для кожної змінної ми можемо вибрати мітку та візуальний канал, який найкраще для неї підійде.

Приклад 1

Країна		Показник
Парагвай		3
Зімбабве		5
Мексика		6
Змінна	Тип даних	Візуальні канали
Країна	категорійні	Орієнтація, Колір, Форма, Текстура, X (або Y)
Показник	кількісні	Орієнтація, Розмір, Значення, X (або Y)

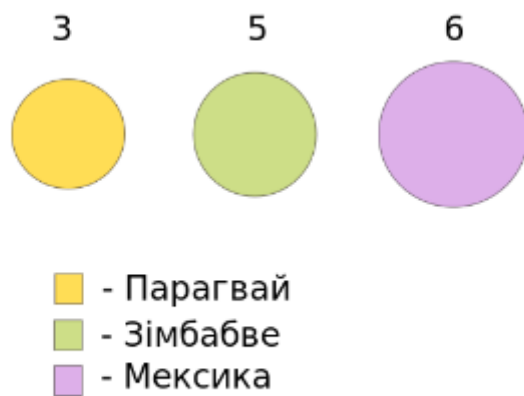
Комбінації:

1. Форма (різні категорії) + Значення (намалюємо число). Картинка, що вийшла виглядає як головоломка, а не як гарна інфографіка.

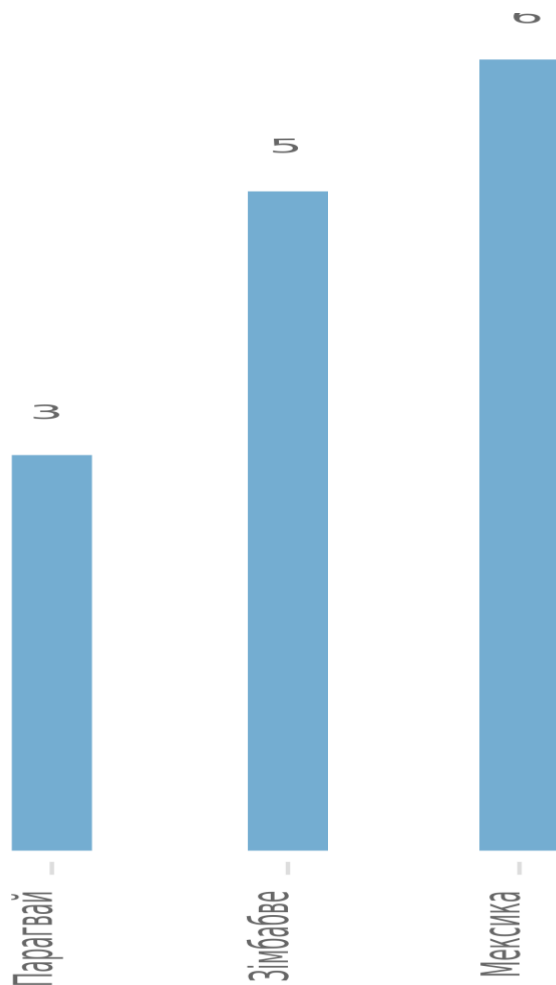


3 - Парагвай
5 - Зімбабве
6 - Мексика

2. Колір (різні країни) + Розмір кола (показники). Виглядає трохи краще, але потрібно вказувати легенду – до якого кольору відноситься яка країна.



3. Значення по Y (показник) + по X (країна). Майже ідеально, бо немає легенди. Підписи на графіку значно краще легенди.



Приклад 2. Складніший приклад

Тип задачі	Приоритет	Кількість зусиль, бали	Час, дні
Характеристика	Обов'язково	30	40
Характеристика	Добре мати	20	40
Характеристика	Непогано мати	15	20
Помилка	Виправити негайно	2	2
Помилка	Виправити колись	2	8
Помилка	Виправити вчасно	5	12
Побажання користувача	Обов'язково	8	10
Побажання користувача	Непогано мати	5	7
Побажання користувача	Добре мати	8	7







Якщо використовувати візуальні канали, які згадувались, то виходить більше 500 варіантів комбінацій. Я виберу лише ті, які вважаю прийнятними у цьому випадку:

Змінна	Тип даних	Візуальний канал
Тип задачі	категорійні	відтінок кольору
Приоритет	впорядковані	насиченість кольору
Кількість зусиль	кількісні	X
Час	кількісні	Y



Після наших прикладів, ви напевно вже зрозуміли, що не всі візуальні канали можна використовувати для всіх типів даних. Для початку, давайте розіб'єм всі візуальні канали на дві частини:

- ті які можна використовувати для кодування категорійних (у першому випадку),
- або кількісних (та впорядкованих) даних (у другому випадку).

	категорій	впорядко	кількісні
	X	X	X
	X		
	X		
		X	X
		X	X
	X		

Канали для відповіді на питання "Що?" (категорійні дані)

- форма;
- місце розташування;
- колір (відтінок).

Канали для відповіді "Наскільки багато?" (кількісні), або наскільки сильно? (впорядковані)

- довжина (1D);
- площа (2D);
- об'єм (3D) ;
- нахил;
- позиція;

- колір (яскравість).

Рекомендації щодо перегляду візуалізації або інфографіки:

- ✓ розглядайте їх з двох боків – дивіться, які дані використані, до якого типу даних відноситься кожна з змінних;
- ✓ порівняйте які канали візуалізації були задіяні для того, щоб візуально представити цей конкретний тип даних.

Табличні дані і графіки

Графіки є одним із головних компонент візуалізації. Сьогодні ми навчимося робити декілька простих типів графіків за допомогою онлайн інструменту під назвою [ChartBuilder](#), який створили у редакції видання [Quartz](#).

Однак перед тим як почати, поговоримо про дані, без яких графіки неможливі. Будь ласка, запам'ятайте наступну пораду: завжди зберігайте ваші дані у файлі, для того, щоб у майбутньому можна було до них повернутися, доповнити або виправити, і швидко перебудувати графік. Якщо ви втратите початкові дані, вам прийдеся кожного разу будувати графік з нуля.

Моя рекомендація - використовуйте для зберігання даних електронні таблиці з вашого офісного пакету, або GoogleSheets - щоб дані були доступними онлайн, і зберігайте їх у форматі CSV - простий формат, в якому у кожній строці значення різних змінних відокремлені комами. Для просунутих користувачів я пропоную подивитися на [CSVKit](#).

Отже, ви знайшли дані, записали їх у електронну таблицю. Записуйте дані наступним чином - по горизонталі, в рядках ми записуємо повний набір значень для всіх змінних, що є в наших даних (наприклад, назва, кількість населення, ВВП на душу, інфляція за останній рік, тощо для якоїсь конкретної країни). По вертикалі в колонках, йдуть значення кожної конкретної змінної із наших даних, наприклад інфляція для всіх країн.

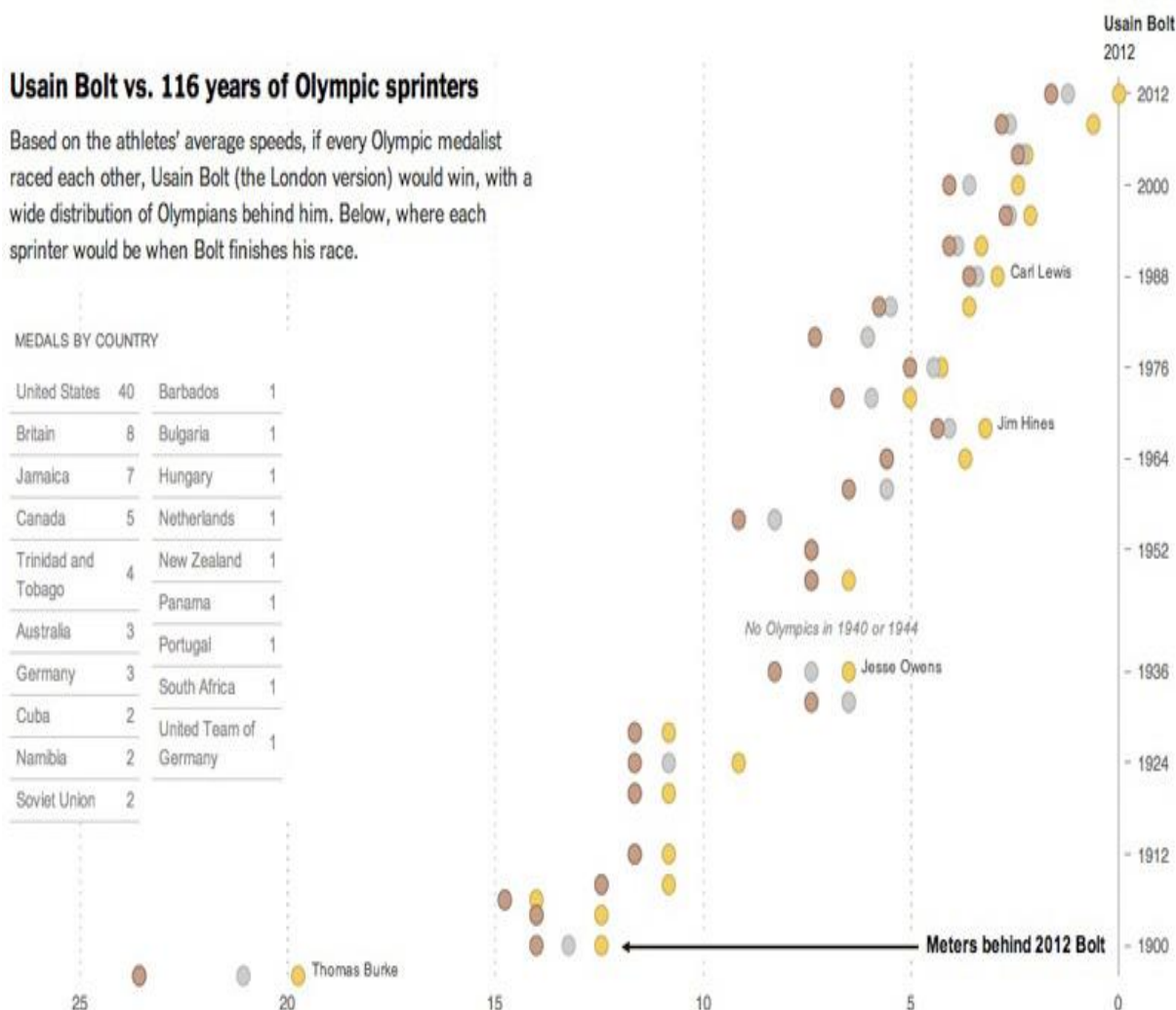
Однак перед тим, як безпосередньо будувати графік, потрібно впевнитися, що дані мають коректну форму. Проста перевірка:

- кожна колонка повинна містити значення лише одної змінної з ваших даних.

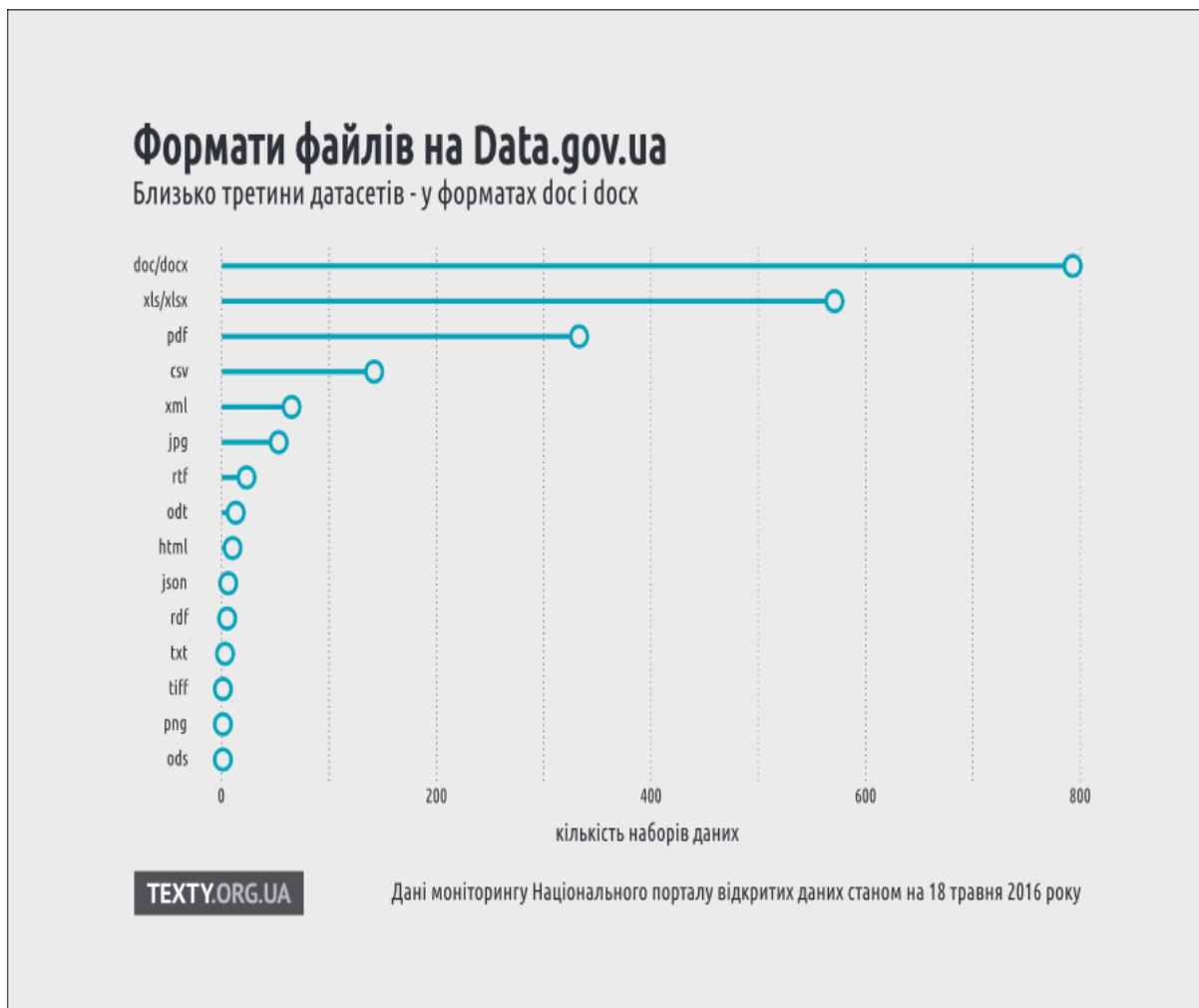
- кількість колонок повинна бути фіксована і однакова для всього файлу (колонки/змінні не з'являються і не зникають, комірки не можна роздвоювати)
- в кожній колонці тип даних має бути однаковим (якщо числа - то всі числа, якщо текст - то весь час текст)
- формат для чисел повинен підходити під інструмент для побудови графіків - наприклад, для ChartBuilder потрібно використовувати точку у якості роздільного знака між цілою та дробною частиною, а не кому.

Аналіз прикладів візуалізацій

Приклад 1. Усейн Болт. Змінні: **роки** (Y), в які були встановлені світові рекорди на дистанцію 100 м, **колір** – тип медалей, **відстань** попередніх спортсменів від Усейн Болта (вісь X)



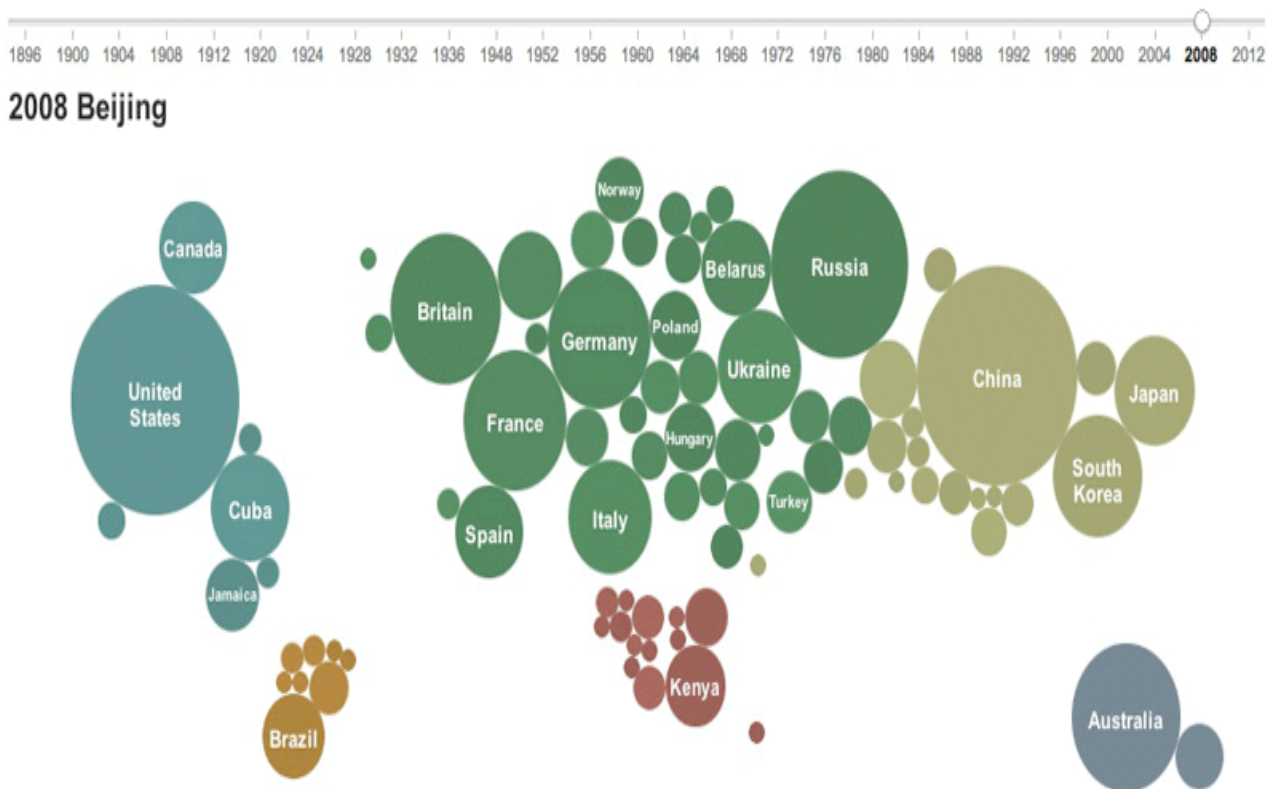
Приклад 2. Графік типу "чупа-чупс". Змінні: Y - тип файлу (категорійні дані), X - кількість цих файлів (кількісні дані)



Позиція є найбільш ефективним способом візуального кодування. Тому графік «Результати олімпіади» – добре показує категорійні дані і кількісні показники, що з ним пов’язані.

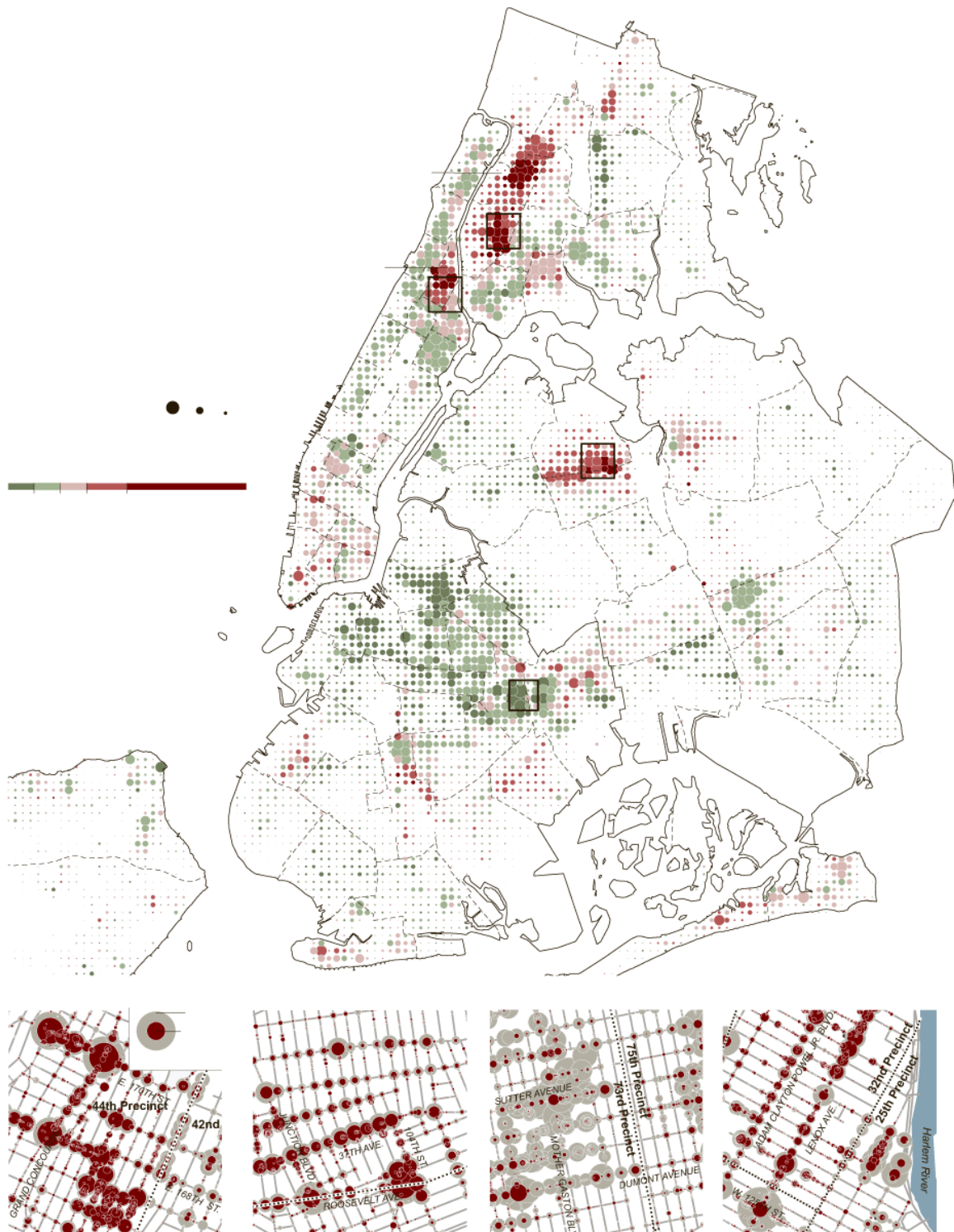
Візуалізація

Візуалізація



Приклад 3. New York, «stop and frisk» - це точкова карта.

Змінні: 1 та 2 - позиції X та Y: зайняті географічними координатами (кількісні дані), 3) кількість обшуків та оглядів, які відбувались у тому чи іншому районі – кількість показана за допомогою розміру кола, та 4 змінна – це інтенсивність обшуку (впорядковані дані) – показана у виді кольору.



Висновки

Сьогодні ми познайомилися з елементами мови візуалізації, з мітками (позначками), та візуальними каналами. Також ми поговорили про те, які типи даних існують, і які візуальні канали більше пасують для різних типів даних.