

Вибір інструментів для візуалізації даних

План

1. Як розповісти історію за допомогою даних
2. Більше, ніж числа
3. Журналістика
4. Залежності
5. Дизайн

Як розповісти історію за допомогою даних

Подумайте про всі популярні книги по візуалізації, які існують, - про ті, про які ви постійно чуєте на лекціях або про які читаєте в блогах, а ще про ті, які прийшли вам на розум прямо зараз, поки ви пробігали очима по цих рядках. Що у них всіх спільного? Вони всі розповідають цікаві історії. Ймовірно, ці історії повинні були переконати вас у чомусь. Або, можливо, підштовхнути вас зробити щось, познайомити з новою інформацією або змусити переглянути свої упереджені уявлення про реальність. Як би там не було, по-справжньому гарна візуалізація даних - неважливо, велика чи маленька, створена з любові до мистецтва або для презентації - допомагає вам зрозуміти, про що саме здатні розповісти дані.

Більше ніж числа

Погляньмо правді в очі. Дані можуть бути чимось досить прикритим і дратівливим, якщо ви не знаєте, що ви шукаєте, або не знаєте, чи є в них щось таке, що слід шукати в першу чергу. Тоді вони перетворюються на купу цифр і слів, які не мають іншого сенсу, крім свого безпосереднього значення. В цьому і полягає великий сенс статистики та візуалізації - в тому, що вони допомагають побачити, що саме стоїть за всім цим. Пам'ятайте: дані є відображенням реального життя. Це не просто купа якихось чисел. І від вас - статистиків, програмістів, дизайнерів і фахівців в обробці даних - залежить, як розповісти ту чи іншу історію.

Більше ніж числа

Як приклад давайте подивимося на безробіття. Видати середні значення по областях країни неважко, але, як ви самі переконалися, стан справ навіть у межах одної області може сильно коливатися. У двох сусідніх районах ситуація може істотно різнитися. Можливо, серед ваших знайомих є людина, яка втратила роботу в останні кілька років, і, як то кажуть, він для вас не просто статистична одиниця, правильно? Вам не потрібно розповідати історію кожного з людей. І все ж є тонка, але від цього не менш важлива відмінність між збільшенням рівня безробіття на 5 відсотків і декількома сотнями тисяч людей, що залишилися без роботи. У першому випадку сказане сприймається як досить абстрактна цифра, а в другому дані мають зовсім інше смислове навантаження.

Журналістика

Повною мірою все вищесказане до мене дійшло під час стажування в газеті New York Times. Тривала вона всього три літніх місяці після другого курсу магістратури, але змінила мій підхід до даних назавжди. Я не тільки навчився створювати графіку для новин. Я навчився підносити дані як новину, а разом з тим дізнався багато про дизайн, організації та перевірки фактів, а також про їх пошуку і вивченні.

Одного разу я присвятив день завданню перевірки трьох чисел з цілого масиву даних, а все тому, що коли відділ графіки New York Times створює якусь діаграму, він повинен упевнитися в точності відомостей. Тільки переконавшись у надійності отриманої інформації, ми переходили до створення презентації. Саме така увага до деталей робить графіку в цій газеті настільки хорошою.

Журналістика

Подивіться на будь-яку схему або діаграму у New York Times. Дані в ній завжди представлені коротко, ясно і надзвичайно красиво. Однак що це означає? Коли ви дивитеся на інфографіку, ви отримуєте можливість зрозуміти дані. Важливі моменти або області супроводжуються коментарями; символи і кольори акуратно пояснені в легенді або в примітках; а ще Times робить все для того, щоб історія, що міститься в даних, прочитувалась аудиторією легко. Це вже не просто діаграма або схема. Це графіки.

Графік, представлена на наступному слайді, схожий на той, яку ви можете знайти в New York Times. На ній показано, як збільшується ймовірність того, що ви помрете в найближчий рік, в залежності від вашого віку.

Вероятность смерти

С возрастом вероятность того, что вы умрете в течение года, увеличивается



Журналістика

В основі цього малюнка лежить простий лінійний графік. Однак елементи дизайну допомагають краще розповісти історію. Підписи і примітки пояснюють контекст і допомагають зрозуміти, чому ці дані цікаві; а товщина ліній і колір фокусують очі на те, що важливо.

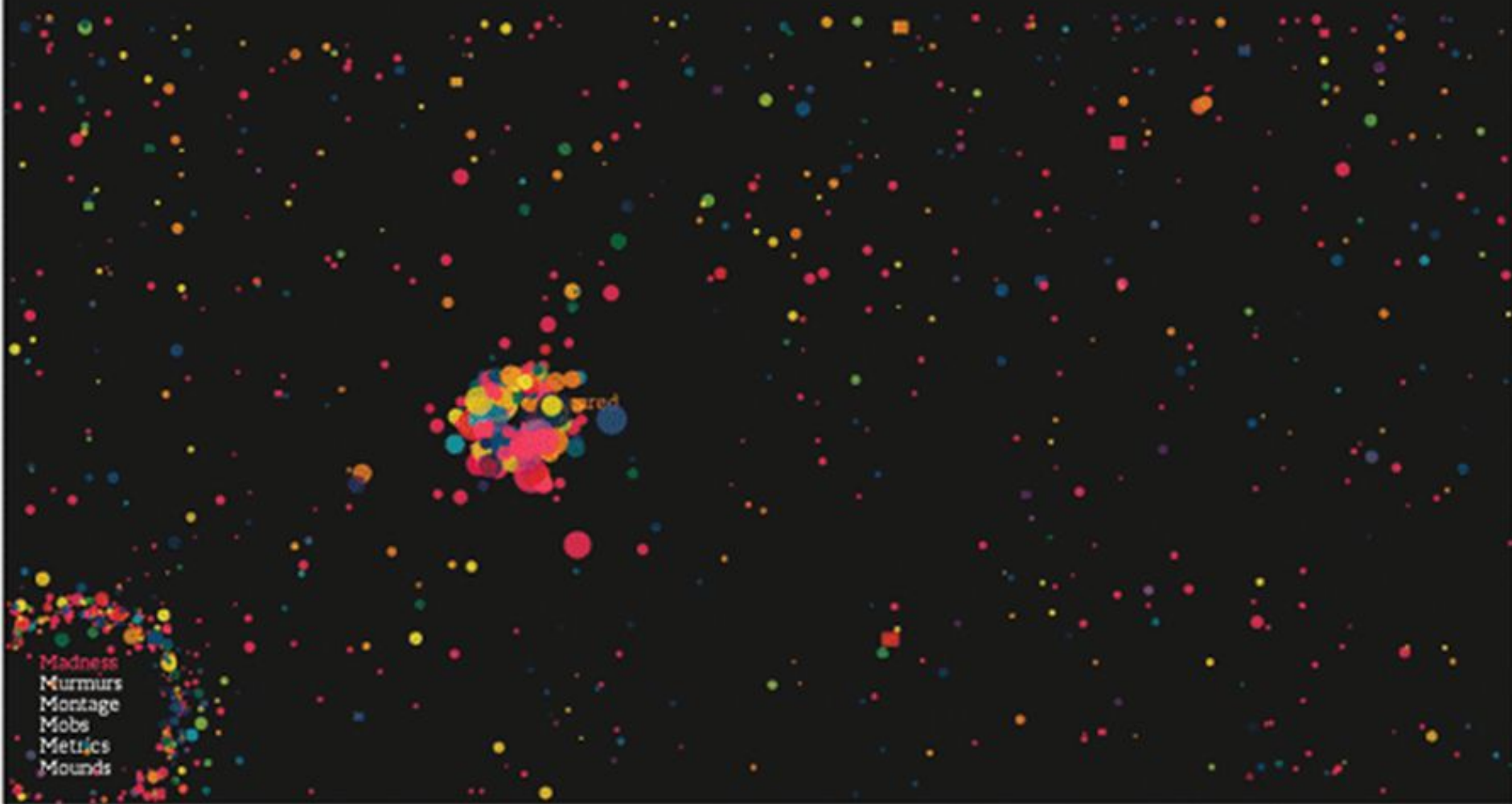
Вимоги до дизайнера графіків і діаграм не зводяться тільки до візуалізації статистичних даних - від нього вимагається також дати зрозуміти, що саме виражає конкретний зображений елемент.

Мистецтво

New York Times - газета об'єктивна. Вона викладає вам дані і дає факти. У цьому їй немає рівних. На іншому кінці спектру знаходиться візуалізація, яка в меншій мірі пов'язана з аналітикою і в більшій - з зануренням в емоції. Саме цим займаються Джонатан Харріс і Сеп Камвар в своїй роботі «Все у нас добре» (Jonathan Harris, Sep Kamvar, We Feel Fine), представленої на наступному слайді.

я думаю о музыке в моей голове но иногда я делаю это пытаюсь скрыть от вас какие чувства я испытываю и вы не узнаете что я подавлен или взбешен или тоскую или нервничаю или что-то еще хотя возможно в этот момент на самом деле у меня в душе клоочет буря эмоций

23 mins ago / from someone




Мистецтво

В їх інтерактивному творінні зібрані пропозиції і фрази з різних персональних блогів, і всі вони візуалізовані як бульбашки, що пливуть в просторі. Кожна бульбашка являє собою емоцію і пофарбована у відповідний колір. Загалом складається враження, ніби всі ці люди мчать в космосі, але якщо ви подивіться довше, то помітите, як бульбашки починають збиватися разом. Спробуйте впорядкувати і категоризувати їх, використовуючи інтерфейс, і ви побачите, як ці, здавалося б, довільно розкидані крихітні штучки об'єднуються. Натисніть на який-небудь пляшечку, щоб дізнатися чиюсь індивідуальну історію. Це так поетично і одночасно так пізнавально!

Мистецтво

Можна навести ще багато прикладів. Скажімо, робота Голана Левіна «Смітник» (Golan Levin, The Dumpster). У ній досліджуються записи в блогах, в яких йдеться про розставання зі значимою людиною. Або «Sumedicina» Кіма Азендорфа (Kim Asendorf), в якій розповідається вигадана історія чоловіка, тікає від ураженої корупцією організації, але не словами, а схемами і діаграмами. Або скульптури Андреаса Ніколаса Фішера (Andreas Nicolas Fischer), що зображують економічний спад в Сполучених Штатах.



Hey wassup nothing here just bored I'm single Abe broke up with me this week he thinks were better off as friends so yeah Monday we have a basketball game I love basketball its my second favriote sport but yeah. I like this one guy and a ...

... Alex dropped me and Em home too which was nice, so up until then a great evening, which all went a bit wrong to be honest, as Tom broke up with me. Feel like absolute shit to be honest, but aboslutely adore my friends ...

... u</3 ((hint the "d" meaning i use to love you)) yeah i broke up with u ((but u were always the one to break up with me)) im over you trust me when i say that

... on friday I BROKE UP WIT JEREMY 4 those who thi nk he broke up wit me he did nt I DID!!! yea da rea rea son was nt beuz things did nt work out it was really cuz i FOUND OUT im not sayin who but thanx told me ...

Мистецтво

Основна ідея полягає в тому, що дані і візуалізація не обов'язково повинні передавати лише голі, холодні факти. Іноді те, що вам необхідно, це зовсім не аналітичні прозріння. Часом буває корисніше розповісти історію з емоційної точки зору, щоб підштовхнути глядачів до роздумів над наявними даними. Подумайте над цим. Не завжди кіно повинно бути документальним, і не завжди візуалізація повинна бути у вигляді традиційних графіків і діаграм.

Залежності

Вам напевно коли-небудь потрапляли на очі малюнки з цілим набором графіків і діаграм, розташованих як ніби в абсолютно випадковому порядку? Я говорю про малюнки, яким очевидно чогось не вистачає - як ніби графічного дизайнера колись було подумати над даними, і він поспіхом додав їм якусь форму, аби встигнути до крайнього терміну. Дуже часто те, чого так явно не вистачає в подібних випадках, - і є залежність.

У статистиці під залежністю, як правило, мають на увазі кореляцію і причинність. Множинні змінні можуть бути переплетені між собою складними зв'язками. Шоста глава, «Візуалізація залежностей», присвячена саме цим поняттям і тому, як їх візуалізувати.

Залежності

Однак і на більш абстрактному плані, на якому ви не думаєте про рівняння і критерії перевірки гіпотез, ви можете розробляти діаграми і графіки для візуального порівняння і зіставлення величин і розподілів. Як простий приклад можна привести схему проникнення технологій в життя, взяту з «Доповіді про світовий прогрес» і представлену на наступному слайді.

ТЕХНОЛОГИИ

ПОЛЬЗОВАТЕЛИ ИНТЕРНЕТА



АБОНЕНТЫ
МОБИЛЬНОЙ СВЯЗИ

59,74
на 100 жителей

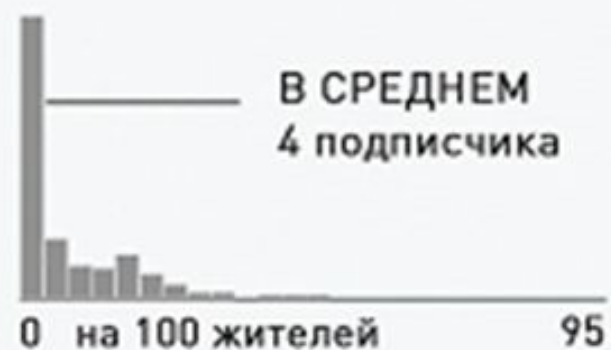
АБОНЕНТЫ С КОММУТИРУЕМЫМ
ДОСТУПОМ



СТАЦИОНАРНЫЕ
И МОБИЛЬНЫЕ СЕТИ

1 к 3,2

ШИРОКОПОЛОСНЫЕ СЕТИ



РОСТ МОБИЛЬНОЙ СВЯЗИ

+285%
с 2003 г.

Залежності

Стовпчикові діаграми показують кількість користувачів інтернету, а також кількість абонентів з комутованим доступом і широкосмугових мереж на 100 жителів. Відзначте, що розкид у інтернет-користувачів (від 0 до 95 на 100 жителів) набагато ширше, ніж у двох інших груп.

Простіше і швидше за все - надати комп'ютерній програмі, яка у вас встановлена, самій вирішити, який діапазон використовувати для кожної столбцовою діаграми. Однак все столбцовою діаграми тут збудовані на базі одного і того ж діапазону, незважаючи на те що немає жодної країни, у якій було б по 95 абонентів з комутованим доступом або користувачів широкосмугових мереж на 100 жителів. Але це дає вам можливість легко порівнювати розподілу між групами.

Залежності

А тому, коли вам доводиться працювати з великою кількістю різних наборів даних, постарайтеся думати про них як про групу, а не як про окремі, жодним чином не пов'язаних між собою масивах. Так ви зможете отримати набагато цікавіші результати

Дизайн

Коли вже всі дані в порядку, ви можете приступати до візуалізації. Над чим би ви не працювали - будь то доповідь, онлайн-інфографіка або твір data-арту, - вам слід дотримуватися кількох основних принципів. Кожен з них дає простір для маневру, так що вам краще сприймати їх, скоріше, як якісь орієнтири, а не як звід жорстких правил, але якщо в графічному дизайні ви новачок, тоді вам саме з них і варто почати.

Дизайн

Розробка будь-якої діаграми протікає приблизно по одному і тому ж сценарію. Ви збираєте дані; ви кодуєте дані за допомогою кіл, стовпців і квітів; а потім пропонуєте іншим людям прочитати їх. В цей момент читачі повинні розшифрувати ваші кодування. Що означають всі ці кола, стовпці і кольору?

В роботі Вільяма Клівленда (William Cleveland) і Роберта МакГі (Robert McGill) питання кодування розглянуті в деталях. Деякі типи кодувань працюють краще, ніж інші. Але який саме підхід до кодування ви виберете, буде не так вже й важливо, якщо в кінцевому рахунку читачі не зможуть зрозуміти, що саме висловлюють ті чи інші позначення. Якщо вони не зможуть розкодувати ваше послання, значить, час, який ви присвятили створенню діаграми, виявиться витраченим даремно.

Дизайн

Можливо, ви відзначали про себе, що в малюнках, що знаходяться десь між data-артом і інфографікою, часом відсутня контекст, а вже з творами data-арту це і зовсім звичайна ситуація. Легенди і мітки здатні повністю зруйнувати враження від роботи, але ви можете принаймні дати трохи інформації, описавши найнеобхідніше в одному короткому абзаці. Це допоможе іншим оцінити вашу працю.

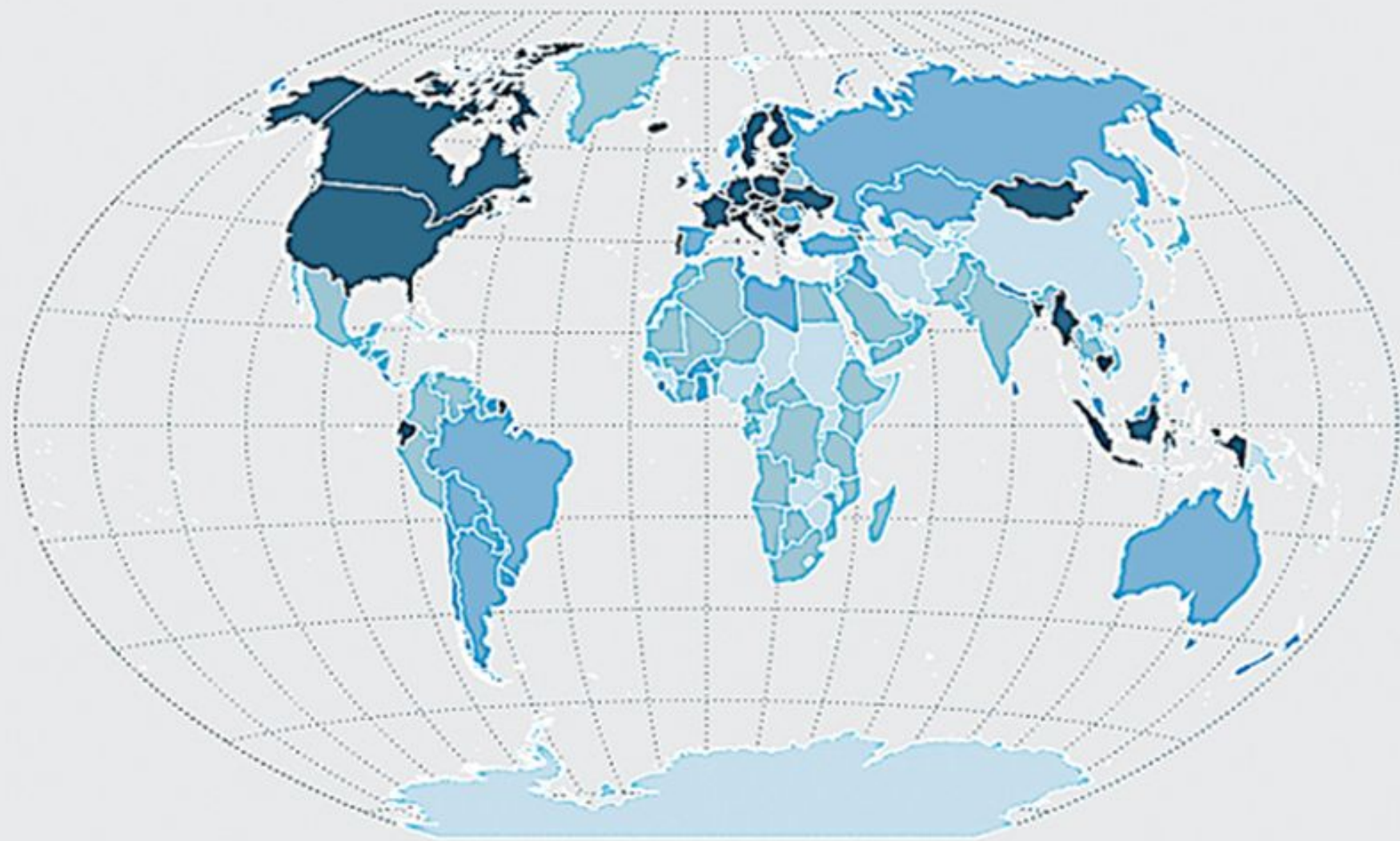
Дизайн

Однак з подібною відсутністю пояснень іноді можна зіткнутися і в тих випадках, коли основне завдання діаграм і графіків якраз і полягає в поданні даних. В результаті читачі бувають розчаровані, а вам напевно не хочеться допускати цього. Відбувається таке деколи тому, що, працюючи з даними, ви з ними зростаються і знаєте напам'ять, що означає кожен елемент діаграми. Читачам жевсе новою. Не знаючи контексту, з яким ви самі звикли, аналізуючи дані, вони виявляються немов сліпими - не можуть побачити те, що бачите ви.

Дизайн

Отже, як ви можете переконатися, що читачам вдасться розшифрувати ваші кодування? Просто поясніть їм значення кожного елемента за допомогою підписів, легенд і ключів. Що саме вибрати - залежить від ситуації. Подивіться, наприклад, на карту світу на наступному слайді, яка дає уявлення про використання програми Firefox в різних країнах земної кулі.

Пользователи Firefox в мире

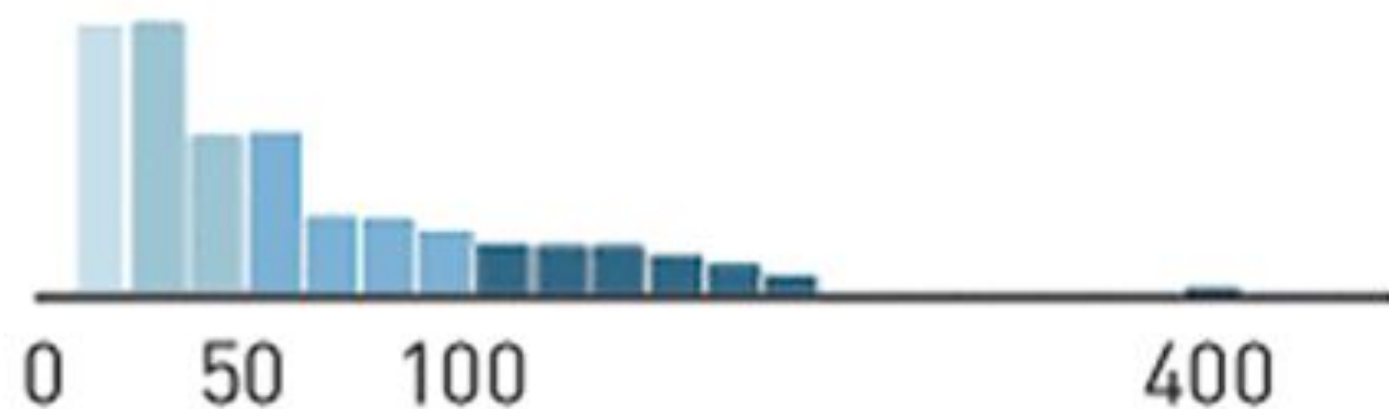


Sources: Mozilla, Internet World Stats

Дизайн

Ви бачите, що різні країни пофарбовані в різні відтінки голубого, але що означають ці відтінки? Темний колір - це більше або менше споживачів? І якщо більше, то більше якого саме кількості? У тому вигляді, в якому карта представлена на малюнку, вона досить-таки марна. Але якщо ви забезпечите її описом, такий, як на наступному слайді, все відразу проясниться. В даному випадку колірна легенда виконує і ще одну функцію: вона являє собою стовпчикову діаграму, яка демонструє поширеність програми на підставі кількості користувачів.

*Количество пользователей Firefox
на 1000 пользователей интернета*

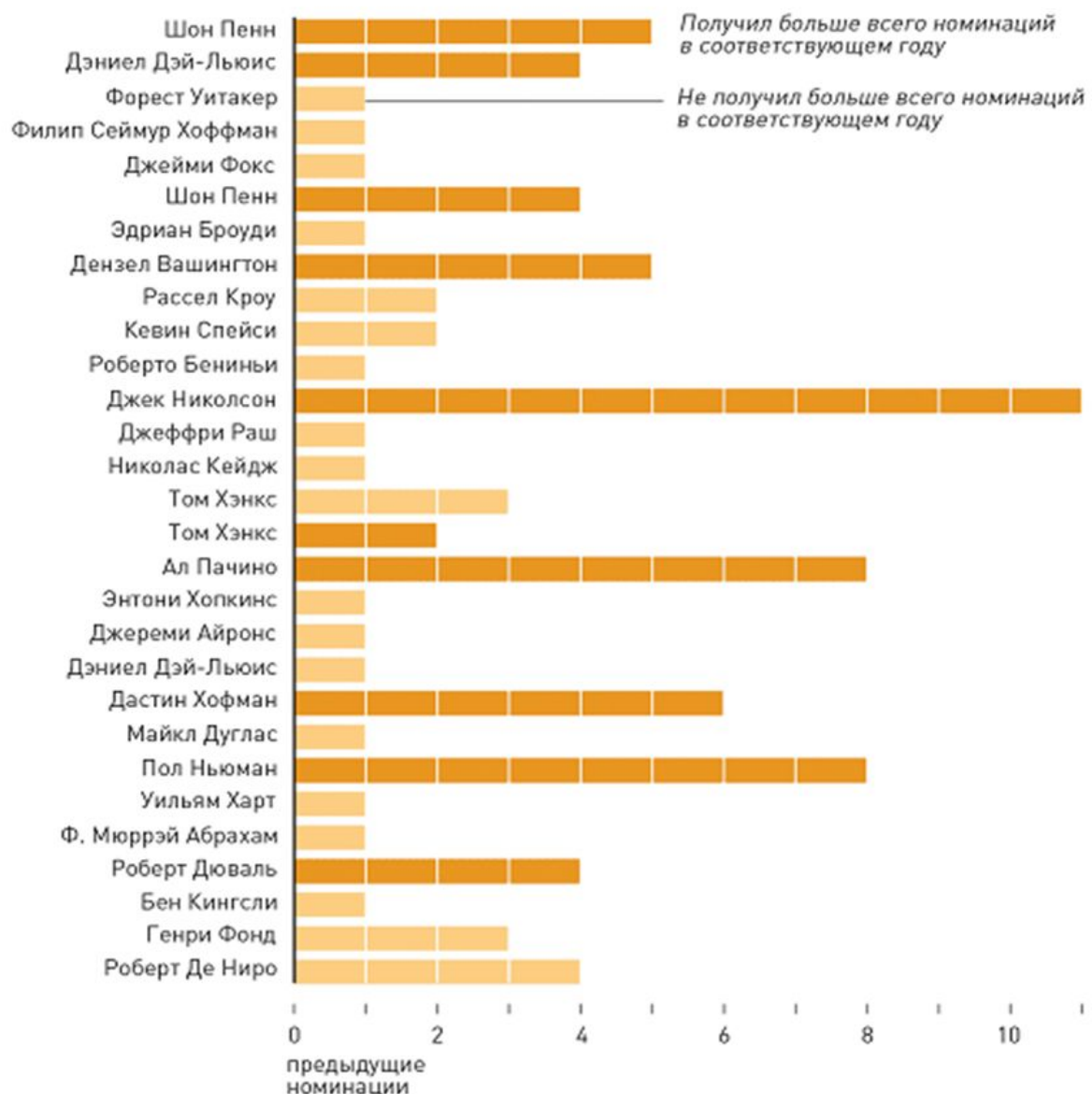


Дизайн

Якщо у вас є достатньо місця або занадто багато категорій, ви можете вставити підписи безпосередньо поруч з об'єктами вашої діаграми, як це продемонстровано на наступному слайді. На даній діаграмі показано кількість номінацій, які той чи інший актор отримав перед тим, як був удостоєний премії «Оскар» за кращу чоловічу роль.

Актеры, получившие «Оскар» за лучшую мужскую роль, и количество их номинаций

Можно ли сказать, что у актеров, имеющих большее количество номинаций на получение премии, также и больше шансов на победу? Только 10 из последних 29 актеров-лауреатов в данной категории номинировались большее количество раз, чем другие претенденты



Дизайн

По інтернету в свій час розгулювала теорія про те, що статуетку, як правило, виграє той актор, який на відповідний рік отримав найбільше номінацій серед представників своєї когорти. Як впливає з підписів, оскарноосці, відмічені темно-оранжевим, дійсно отримали найбільше номінацій, а світло-помаранчевим відзначені ті, хто не отримав найбільше номінацій, однак тим не менше був удостоєний статуетки у відповідному році.

Дизайн

Як бачите, у вас є безліч варіантів дії. Всі ці деталі не вимагають великих зусиль у використанні, вони начебто дрібні, але різниця від їх застосування може бути величезною для сприйняття вашої роботи. По інтернету в свій час розгулювала теорія про те, що статуетку, як правило, виграє той актор, який на відповідний рік отримав найбільше номінацій серед представників своєї когорти. Як впливає з підписів, оскарносі, відмічені темно-оранжевим, дійсно отримали найбільше номінацій, а світло-помаранчевим відзначені ті, хто не отримав найбільше номінацій, однак тим не менше був удостоєний статуетки у відповідному році.

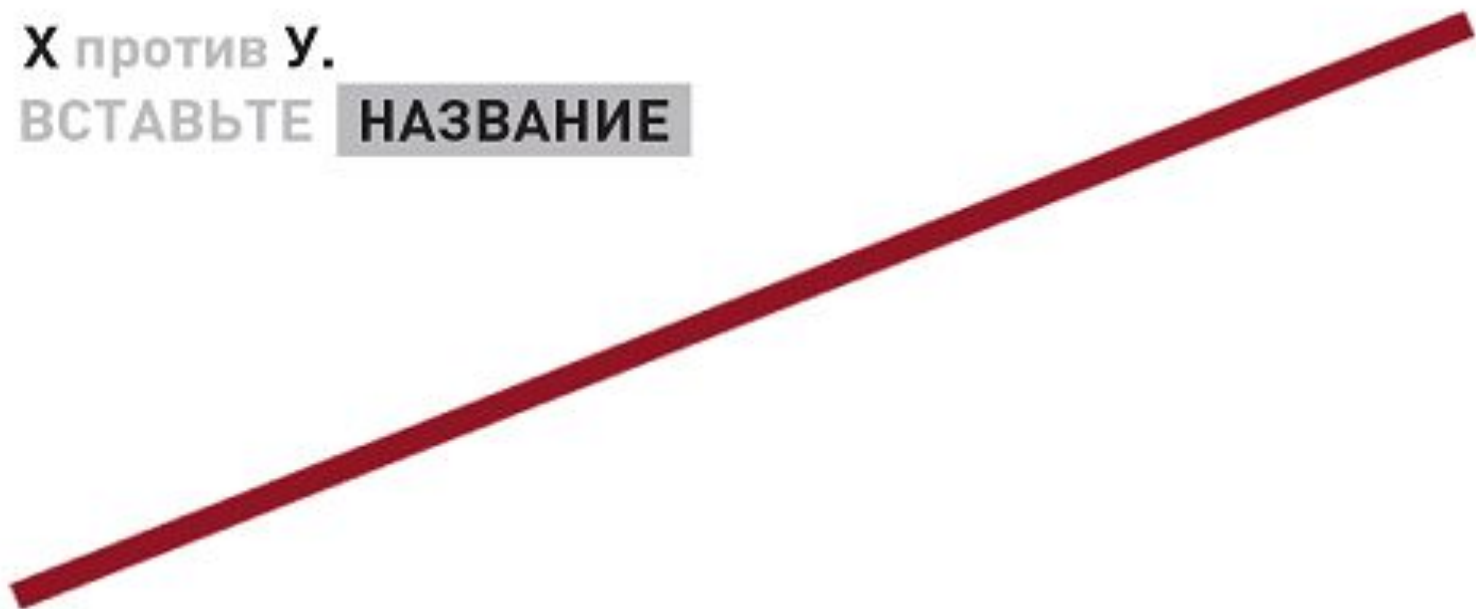
Дизайн

Дайте осям назви

Ви вже зрозуміли, чому необхідно завжди давати пояснення до кодування. Точно так же вам слід завжди давати назви осях в ваших графіках і діаграмах. Без підписів і пояснень осі виконуватимуть лише декоративну роль. Дайте їм назви, щоб читачі знали, яка саме шкала застосовується в кожному конкретному випадку. Логарифмічна, диференціальна, експоненціальна.

X против Y.

ВСТАВЬТЕ **НАЗВАНИЕ**



Дизайн

Як ви можете перекоонатися, притому що графік всюди один, в результаті простої зміни назв осей виходять абсолютно різні історії. Звичайно, все це робилося заради забави. А тепер уявіть собі, що даний графік повинен був розповісти про щось дуже серйозне. Без підписів і назв ваша діаграма просто безглузда.

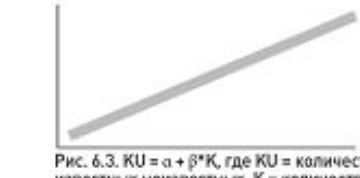
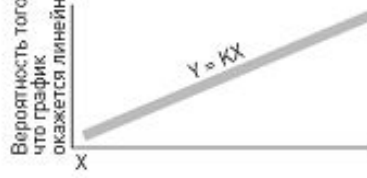
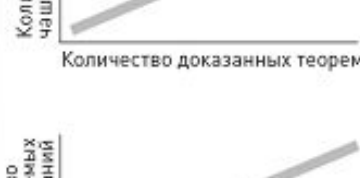
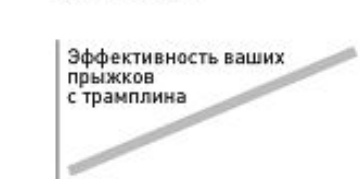


Рис. 6.3. $KU = a + \beta * K$, где KU = количеству известных неизвестных, K = количеству известных, $\beta > 0$ ($p < 0,001$) (используется с разрешения; исходные данные предоставлены «Журналом статистической эпистемологии Среднего Запада»)



Дякую за увагу!