Interrogate Your Data Like a Journalist

Using R to Gather, Vet and Present Data in Journalism

Molly Huie

Team Lead – Data, Legal Analytics & Business (DataLAB)

Andrew Wallender

Data Reporter – INDG Central News

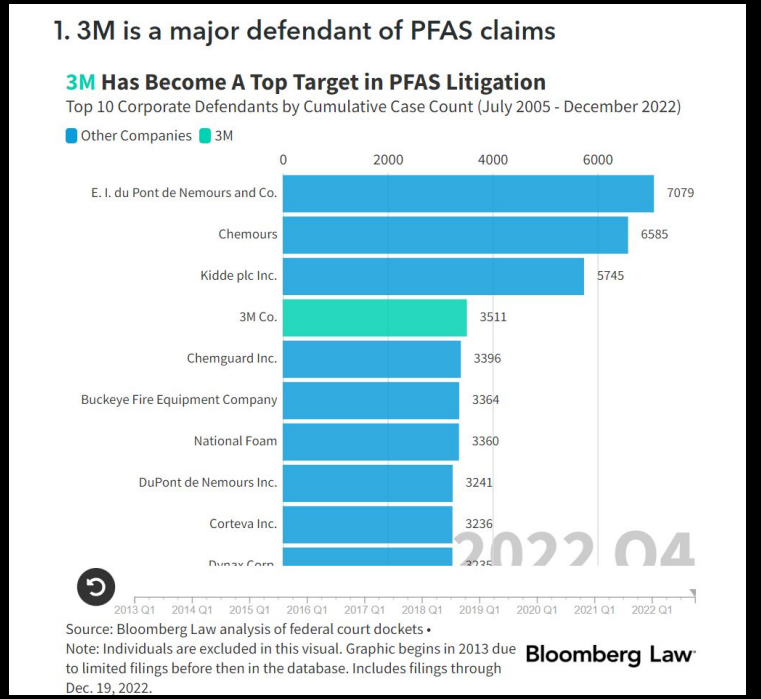
Bloomberg Law

End to End Data Journalism

Go from this…

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE docket SYSTEM "super_docket.dtd">
<docket>
   <docket_body>
      <docket_header>
         <line_list>
            <line>U.S. District Court</line>
            <line>District of South Carolina (Charleston)</line>
            <line>CIVIL DOCKET FOR CASE #: 2:20-cv-00042-RMG</line>
         </line_list>
      </docket_header>
      <info>
         <assigned_to>Honorable Richard M Gergel</assigned_to>
         <cause>28:1332 Diversity-Personal Injury</cause>
         <date_filed>01/03/2020</date_filed>
         <jurisdiction>Diversity</jurisdiction>
         <jury_demand>Plaintiff</jury_demand>
         <lead_case>2:18-mn-02873-RMG</lead_case>
         <member_cases>
            <line_list>
               <line>2:20-cv-01974-RMG</line>
               <line>2:20-cv-01983-RMG</line>
               <line>2:20-cv-02030-RMG</line>
               <line>2:20-cv-02115-RMG</line>
               <line>2:20-cv-02154-RMG</line>
               <line>2:20-cv-02294-RMG</line>
               <line>2:20-cv-02311-RMG</line>
               <line>2:20-cv-02318-RMG</line>
               <line>2:20-cv-02332-RMG</line>
               <line>2:20-cv-02384-RMG</line>
               <line>2:20-cv-02444-RMG</line>
               <line>2:20-cv-02445-RMG</line>
               <line>2:20-cv-02446-RMG</line>
               <line>2:20-cv-02447-RMG</line>
               <line>2:20-cv-02448-RMG</line>
               <line>2:20-cv-02450-RMG</line>
```

To this…



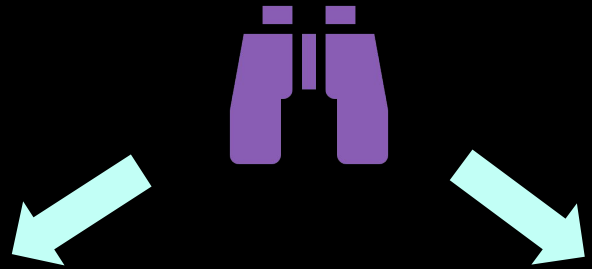
Bloomberg Law

Benefits of Interrogating Your Data
#LikeAJournalist

- **L**ayered Understanding
- **A**pproachable
- **J**ustifiable



Step 1: Gather Data



Internal Data

External / Publicly available Data

Step 2: Vet The Data

- ☑ Format
- ☑ Spot Check
  - Outliers? Missing information?
- ☑ Make No Assumptions

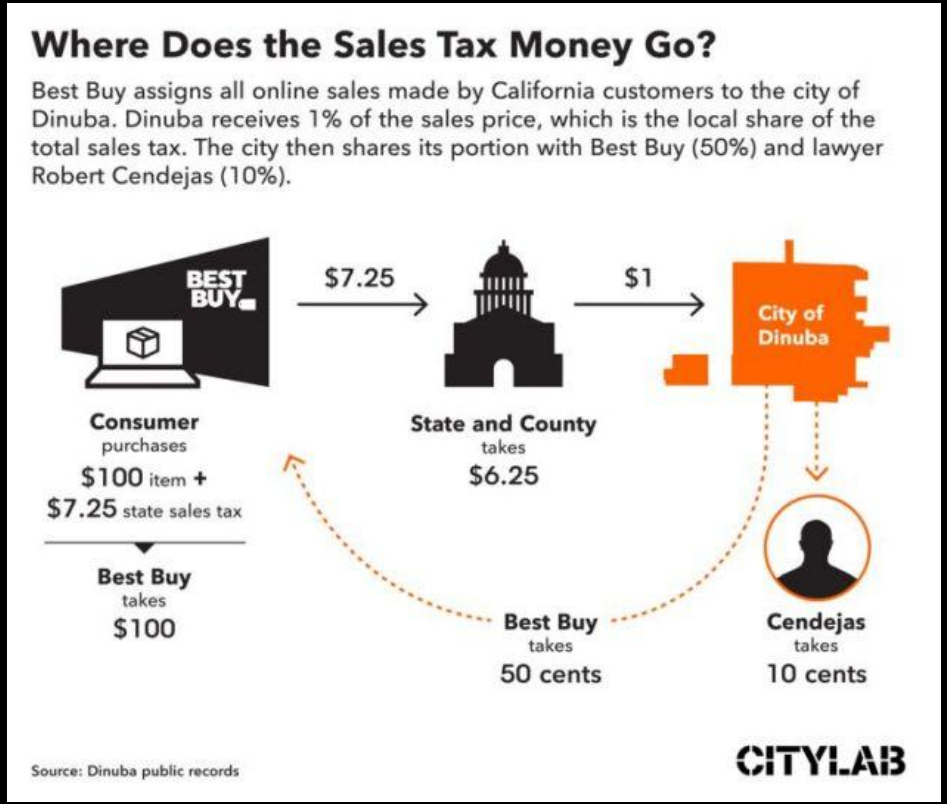
Step 3: Writing About Data

- ☑ Round appropriately
- ☑ Describe in real words
  - ‘About a quarter’ or ‘just over half’
- ☑ Think about seasonality
- ☑ Rates vs percentages

For Example…

Dinuba receives 1% of the sales price, which is the local share of the total sales tax. The city then shares its portion with Best Buy (50%) and lawyer Robert Cendejas (10%).

Talk About Data in Real Terms

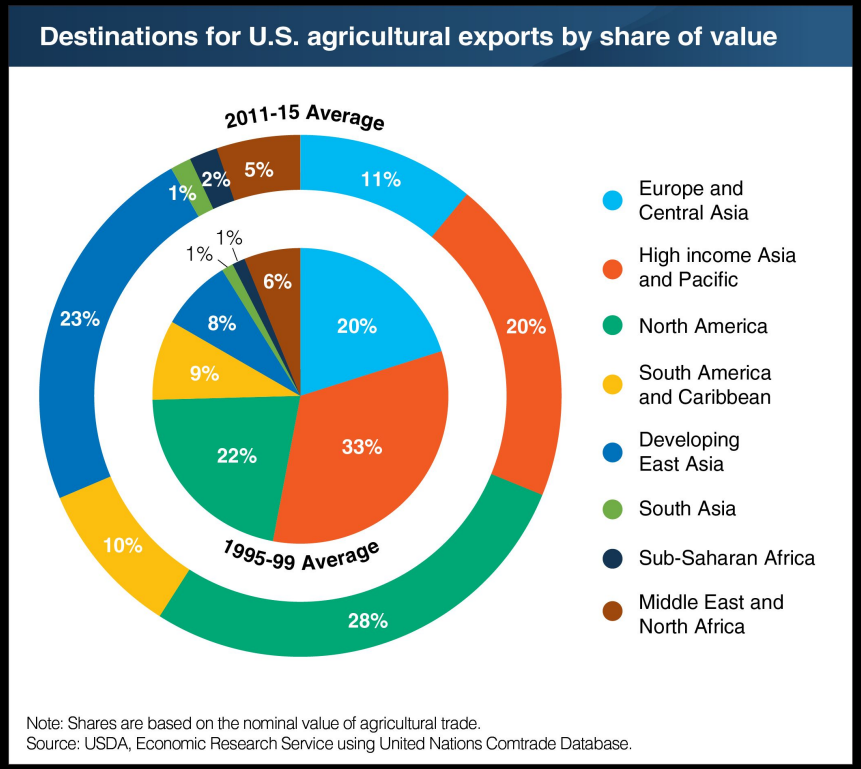


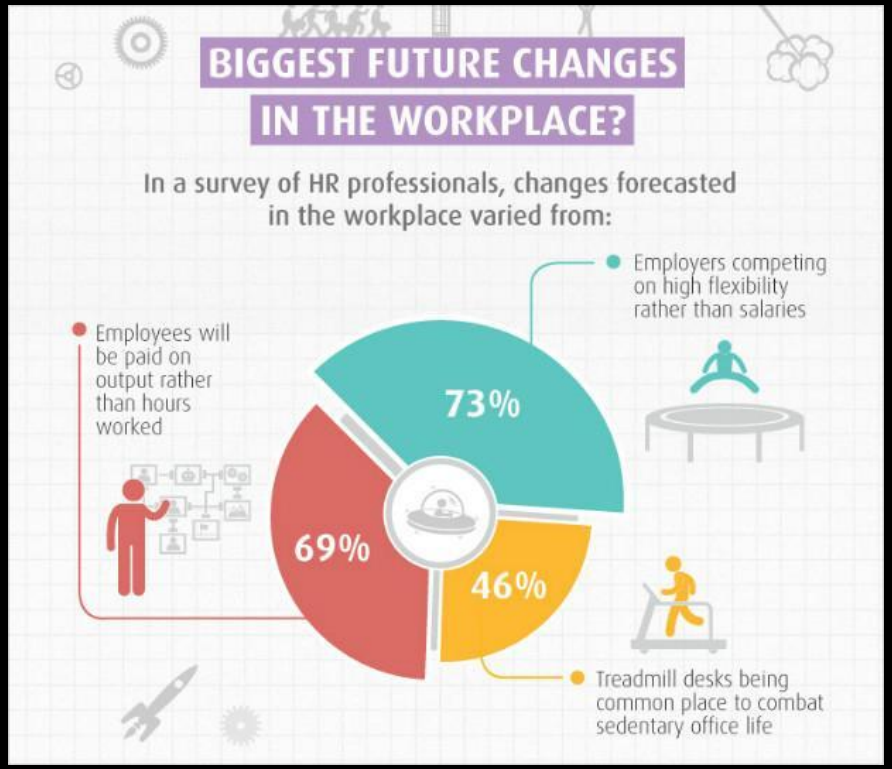
From "Big Tech's California Tax Deals Pit Have-Not Cities Versus Haves" by Laura Mahoney (February 24, 2023)

Step 4: Visualizing the Data

- [x] Don’t be wrong, don’t be confusing
- [x] Use appropriate chart types
- [x] Always include your source
- [x] Visualizations should stand alone

Please don’t do this…..





Bloomberg Law

Journalist Wish List



1. Data Dictionary / detailed explanation of the data fields

2. Clear point of contact for data questions (ideally separate from a press contact)

3. Indication of when the data was last updated and if it's still being updated

Journalist Wish List (Cont.)



4. **Make the data easily downloadable via CSV, XML, JSON, or another standardized format**
   - *Don't* lock the data away behind a dashboard!

5. **Tidy up the data**
   - Multiple spreadsheets are better than a spreadsheet with 70+ columns

Interrogating Your Data
#LikeAJournalist

- Gather
- Vet
- Write
- Visualize

