

**N**

**A**

# N eighborhood A nalysis



# N eighborhood A nalysis



Data  
Visualization

NEVER HAVE I FELT SO  
CLOSE TO ANOTHER SOUL  
AND YET SO HELPLESSLY ALONE  
AS WHEN I GOOGLE AN ERROR  
AND THERE'S ONE RESULT  
A THREAD BY SOMEONE  
WITH THE SAME PROBLEM  
AND NO ANSWER  
LAST POSTED TO IN 2003

WHO WERE YOU,  
DENVERCODER?

WHAT DID YOU SEE?!



# Last Session

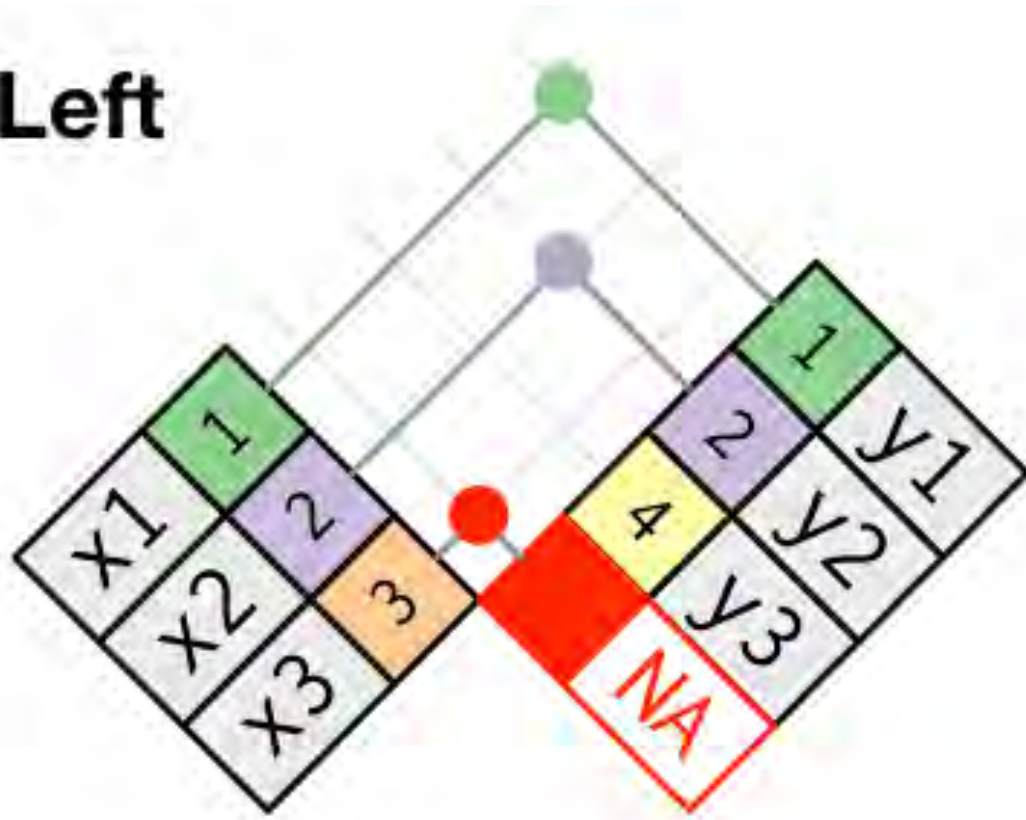
- Relational joins
- Recoding values using `case_when()`
- Developing policy-relevant summaries

# This Week

- Data visualization basics

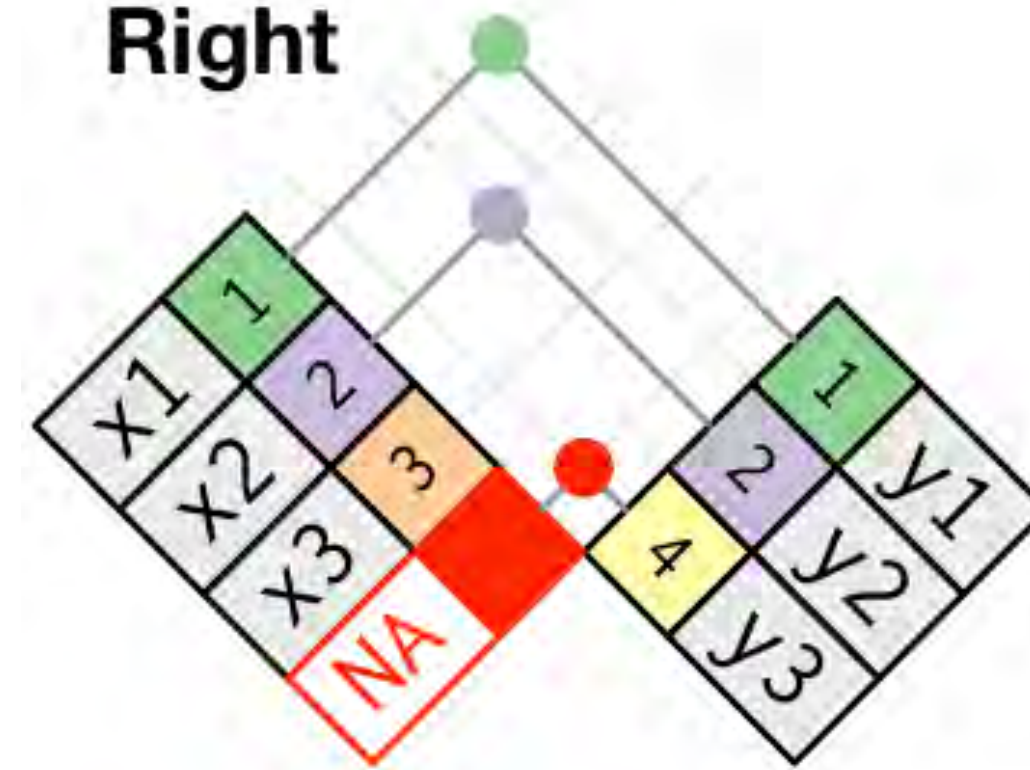
# Debrief - Last Week's Work

Left



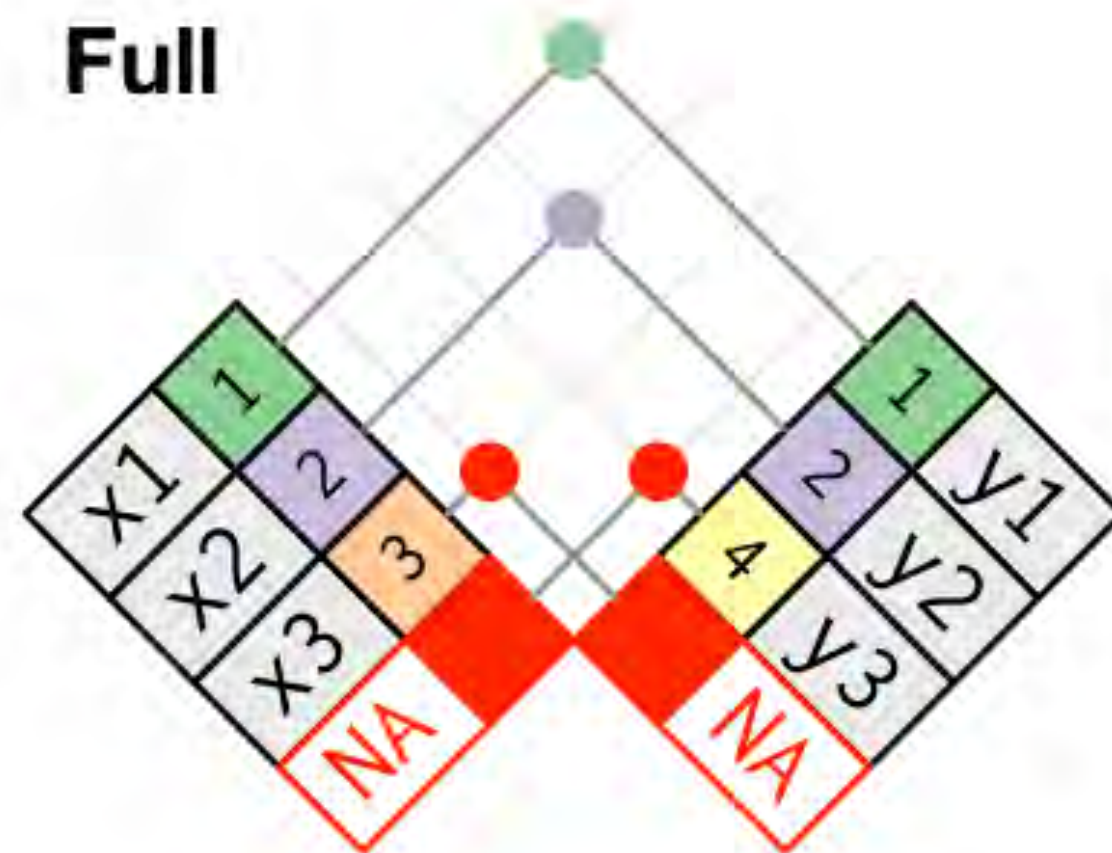
key	val_x	val_y
1	x1	y1
2	x2	y2
3	x3	NA

Right



key	val_x	val_y
1	x1	y1
2	x2	y2
4	NA	y3

Full

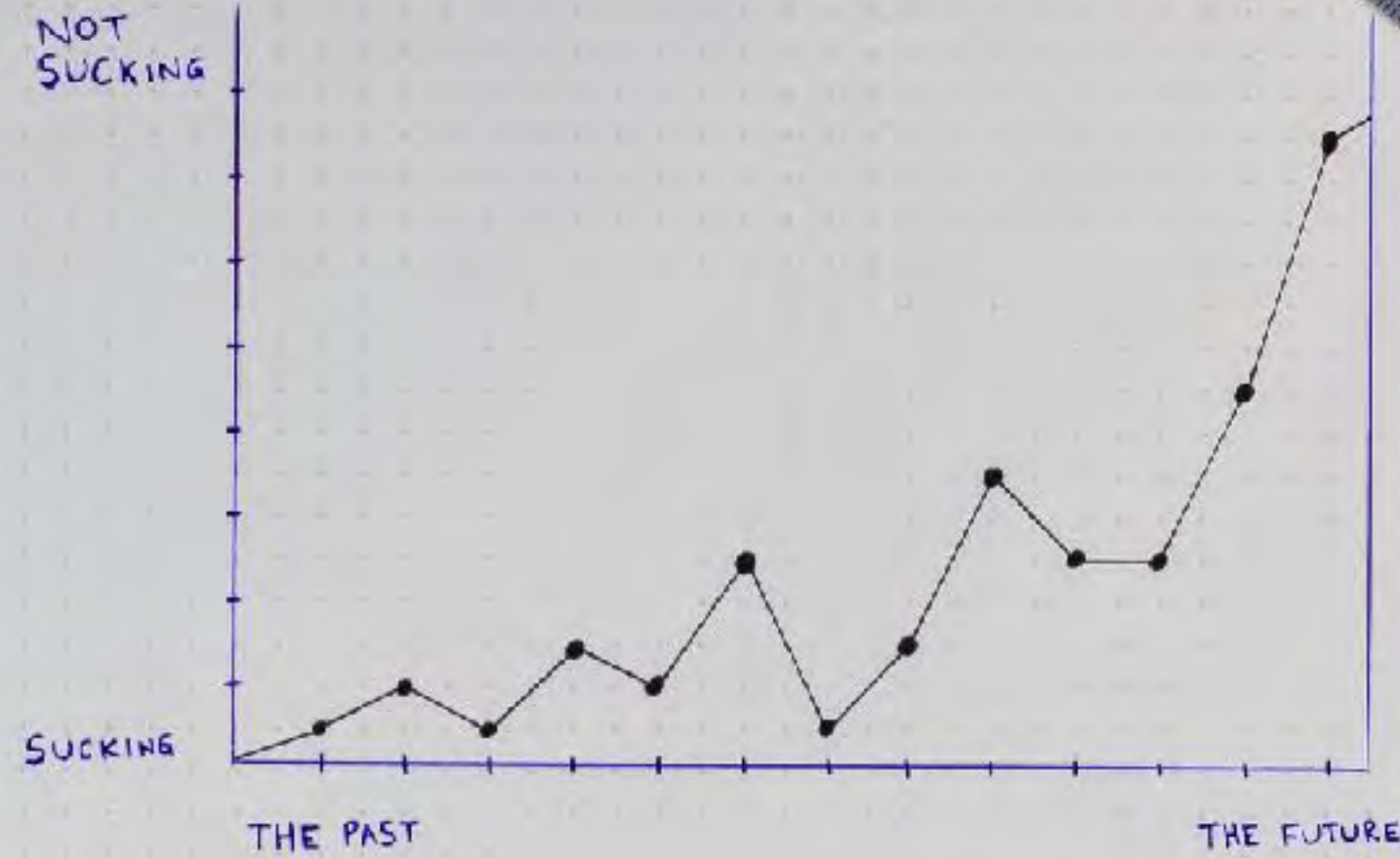


key	val_x	val_y
1	x1	y1
2	x2	y2
3	x3	NA
4	NA	y3

# A Process Note



# Data Visualization





# Why do we visualize data?



## The Georgia Negro



## A Series of Statistical Charts Illustrating the Condition of the Descendants of Former African Slaves Now in Residence in the United States of America



LAND OWNED BY NEGROES IN GEORGIA, U.S.A. 1870-1900.



“I got a couple of my best students and put a series of facts into charts: the size and growth of the Negro American group; its division by age and sex; its distribution, education and occupations; its books and periodicals. We made a most interesting set of drawings, limned on pasteboard cards about a yard square and mounted on a number of moveable standards. The details of finishing these 50 or more charts, in colors, with accuracy, was terribly difficult with little money, limited time and not much encouragement.”

**LAND OWNED BY NEGROES IN GEORGIA, U.S.A. 1870-1900.**

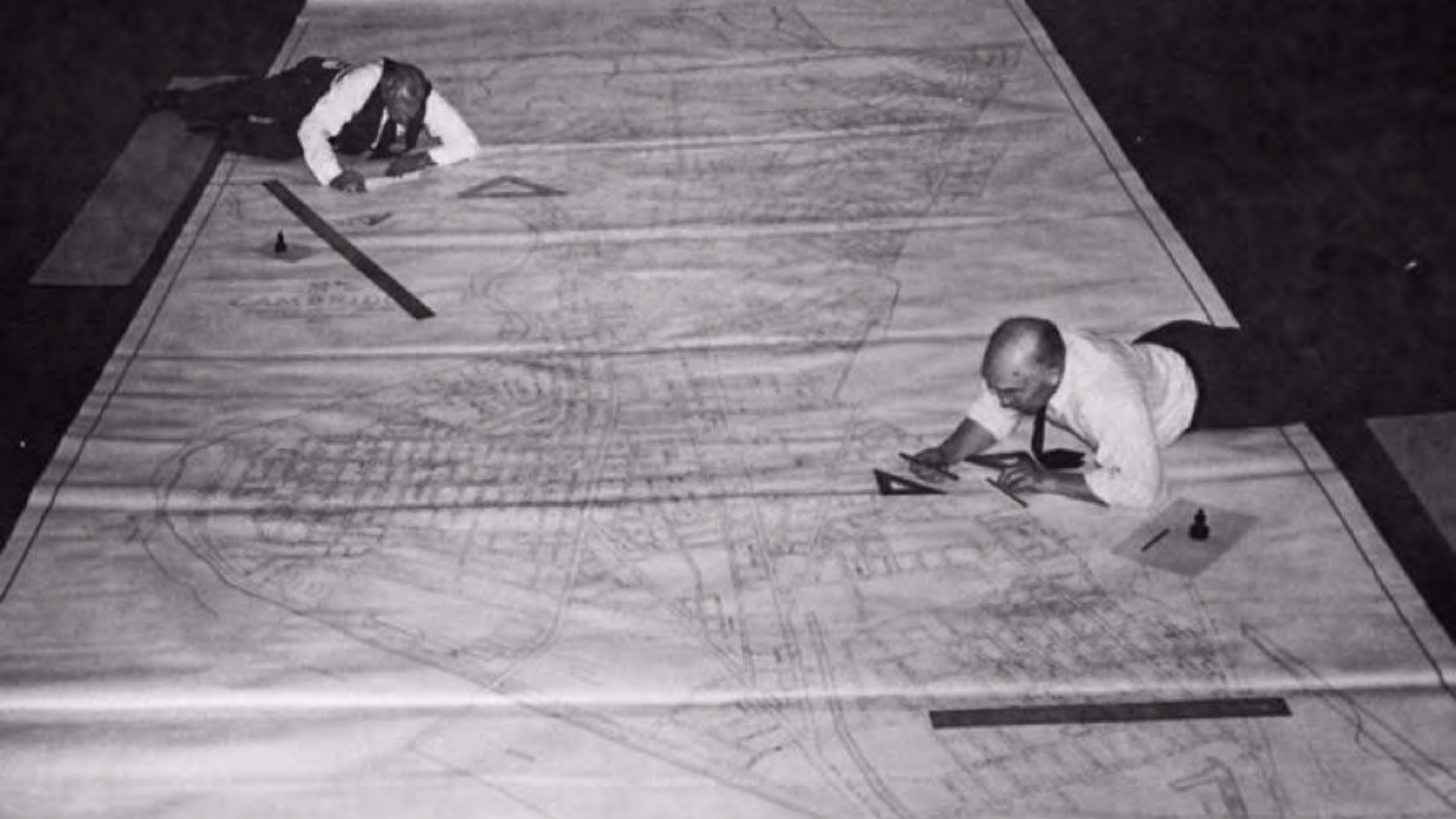


THE FIGURES INDICATE THE NUMBER OF ACRES OWNED IN EACH COUNTY IN 1899.

**CITY AND RURAL POPULATION. 1890.**

7899 NEGROES IN CITIES OF OVER 10,000 INHABITANTS  
 3025 NEGROES IN CITIES FROM 5,000 TO 10,000  
 37689 NEGROES IN CITIES FROM 2,500 TO 5,000







International Business Machines Corp., N. Y. C.

### A. The Use of a Cosmograph to Make a Flow Chart.

1. The "Cosmograph" is a flow chart made by using the device shown above. One thousand strips of paper are set on edge to represent 100%, and are separated into component parts of 100%.
2. These two illustrations give two steps in making a "Cosmograph." The first shows the process of locating and firmly clamping the strips of paper into position. The second shows wedge spacers and bar spacers being inserted between groups of strips of paper.

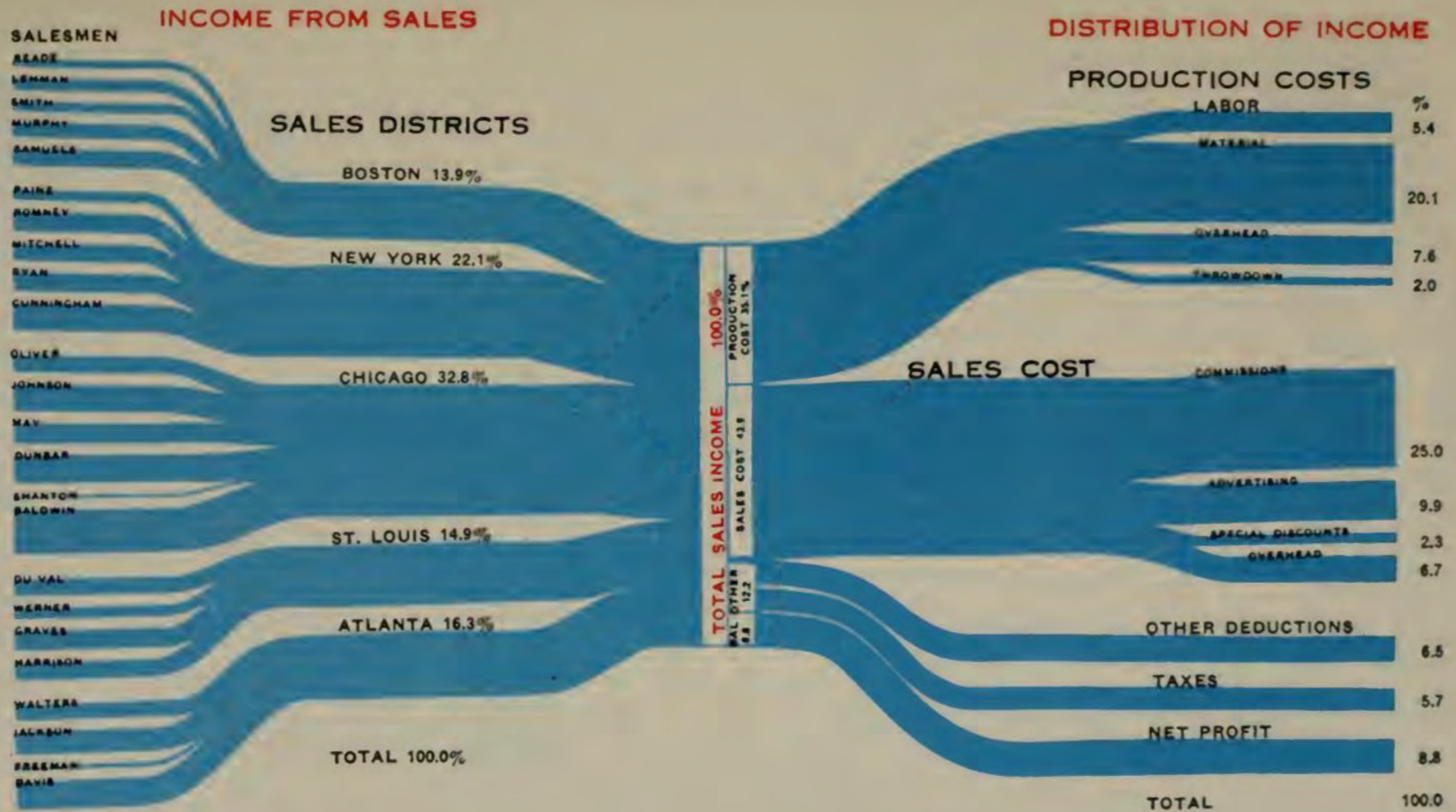


The first or negative photostatic print of the Cosmograph set-up at the left

International Business Machines Corp., N. Y. C.

## B. The Completed Cosmograph.

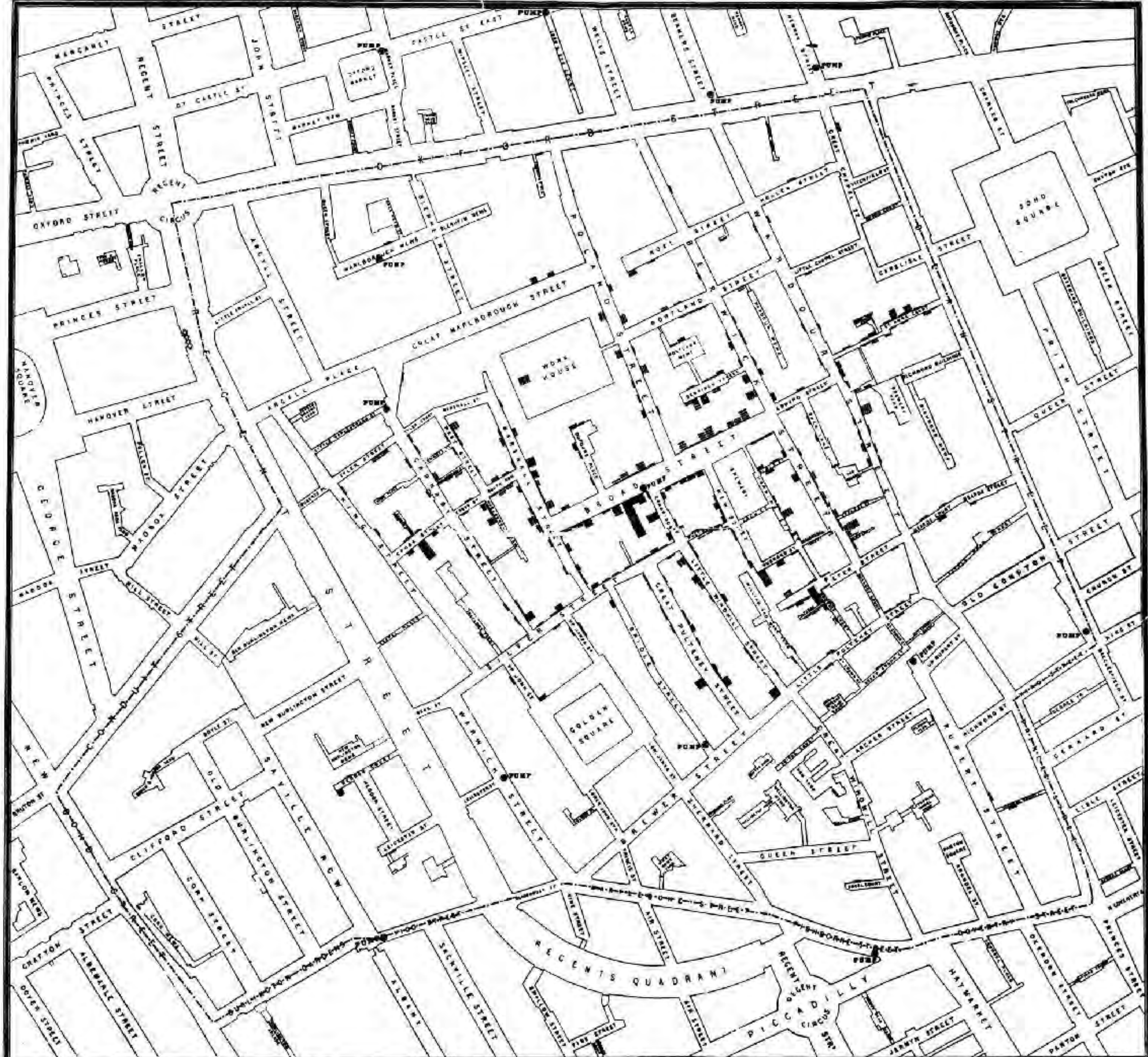
1. Border guides are placed in position to block out excess ends of the paper strips and the Cosmograph is ready for photostatting.
2. The negative photostatic print appears at the right. Note that all black portions of the device fail to reproduce. Of the one thousand strips of paper, twenty are red and are set at each 5% mark. In the negative photostat, these red strips of paper reproduce as white.



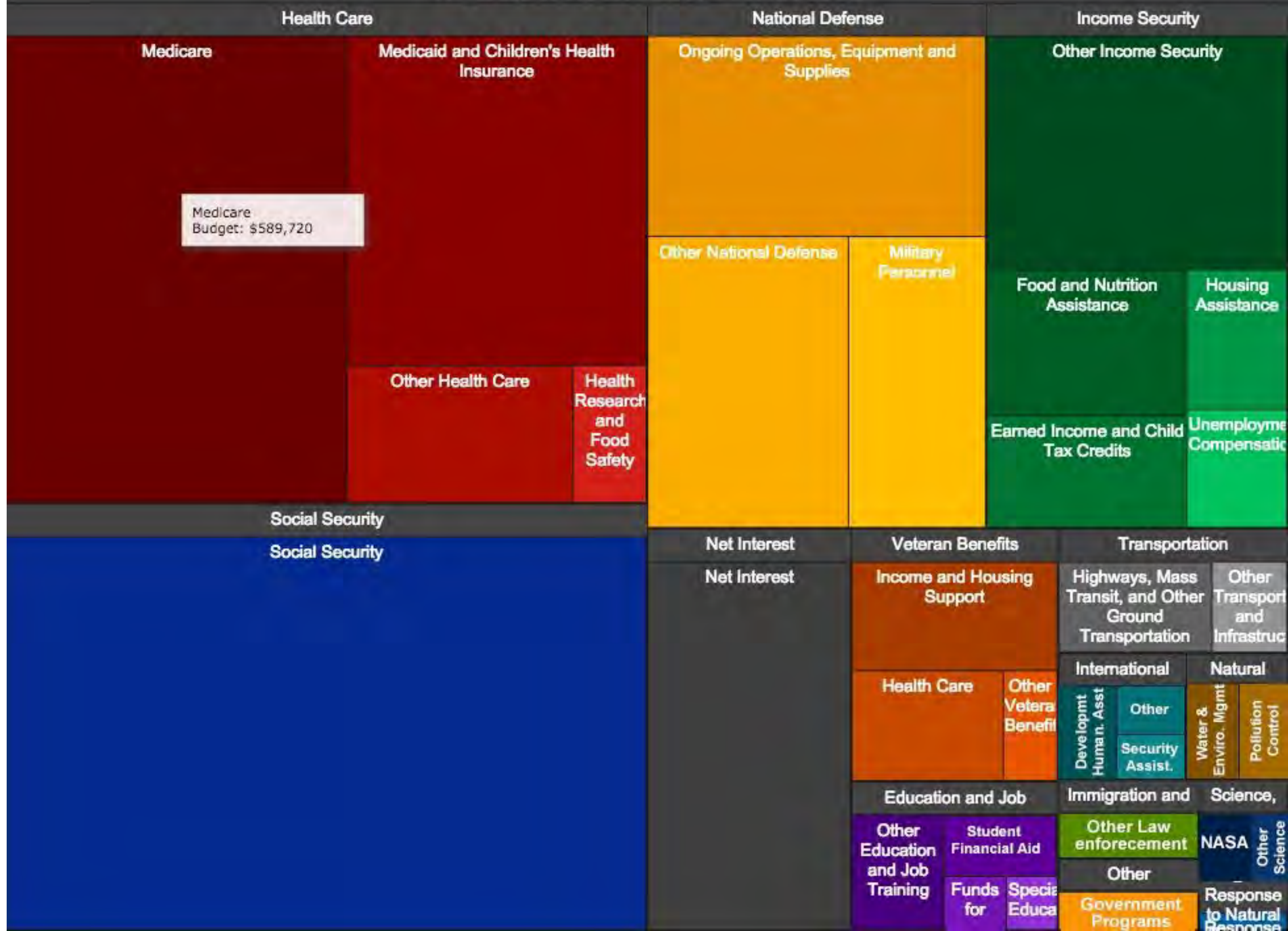
A Cosmograph Showing Distribution of Income From Sales.

The right side of the chart shows the distribution of the income on the left. It shows not only the major channels of distribution —production cost, sales cost, etc.—but also the factors of which these major costs are composed.





2016 outlays from the 2016 Budget, in millions of dollars



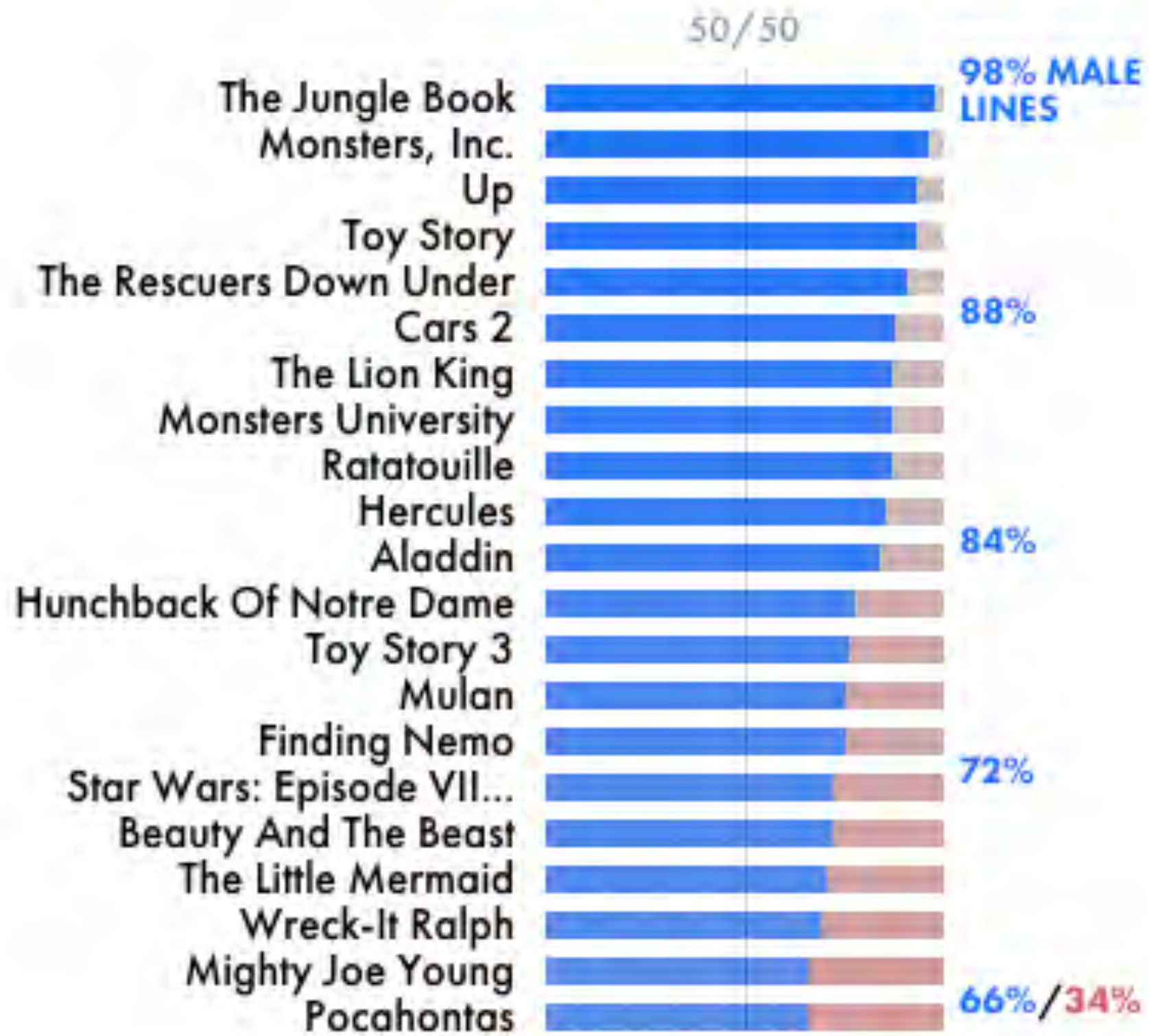


## Screenplay Dialogue, Broken-down by Gender

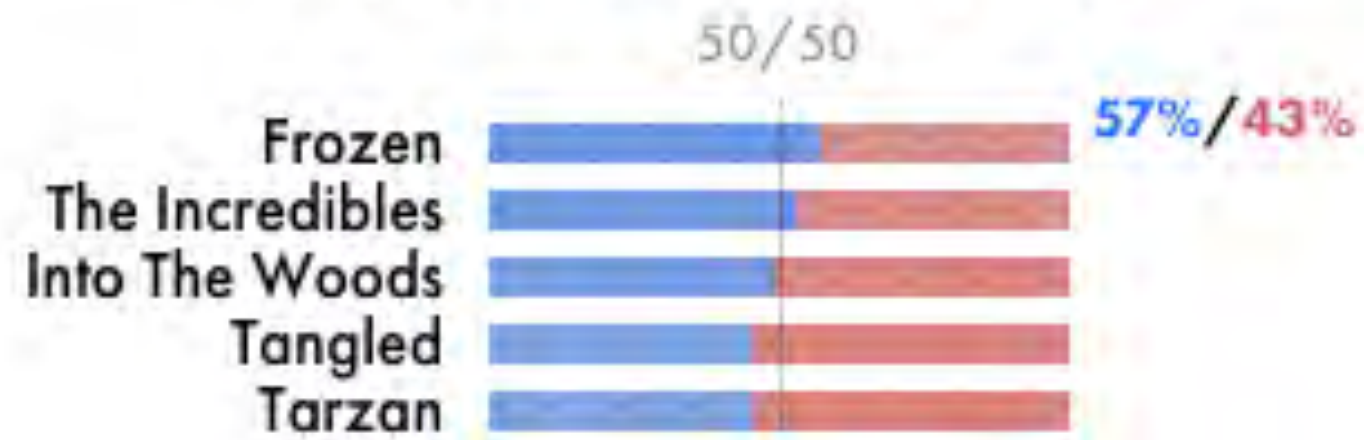
2,005 Screenplays: Dialogue  
Broken-down by Gender

Only High-Grossing Films: Ranked in  
the Top 2,500 by US Box Office\*

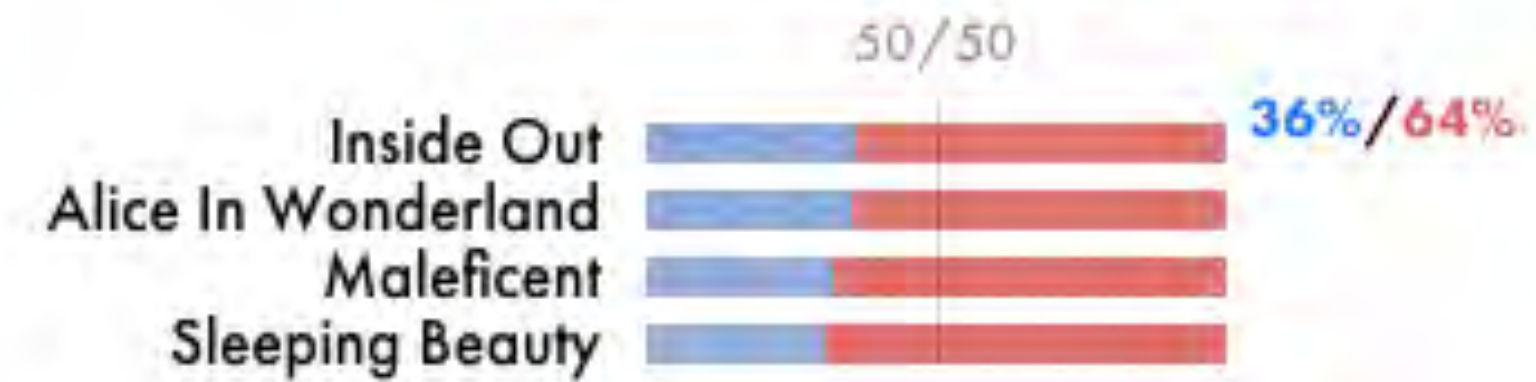
### Men have 60%+ Lines



### Gender Balance, +/- 10%



### Women have 60%+ Lines



What they **should** be spending:

**50% Needs:**  
\$2,241

**30% Wants:**  
\$1,345

**20% Saves:**  
\$896



What they **are actually** spending:

**71% Needs:**  
\$3,189

**17% Wants:**  
\$776

**12% Saves:**  
\$517





2020-01-28



This week we're learning the **grammar of graphics** in GGplot



This week we're learning the **grammar of graphics** in GGplot



**Data:** Your source data

**Aesthetics:** What variables are visualized and how

**Geometry:** The specific plot type we would like to make

**Facets:** Facets allow us to create small multiples

**Statistics:** summaries of data distribution

**Coordinates:** Coordinate system for the plot

**Themes:** Collections of plot styles and options

```
ggplot(duplo, aes(x=index, group =  
color)) +  
geom_bar(aes(fill = color)) #rstats
```





- **ggplot():** make a new plot
  - **duplo:** the data to plot
  - **aes():** aesthetic mapping
    - **x:** what to plot on the x-axis, in this case the variable from duplo called index
    - **group:** in this case a categorical variable called color from Duplo
  - **geom\_bar():** the geometry type to use in the plot
  - **fill():** what to fill the bars with, in this case, using categories from the variable color from the Duplo dataset

```
ggplot(duplo, aes(x=index, group =  
color)) +  
geom_bar(aes(fill = color)) #rstats
```





# Your Lab

- Introduces ggplot

# Questions

