EXTRACTING INFORMATION FROM DATA: VISUALIZING GERONTOLOGICAL RESULTS

TAKA YAMASHITA ¹
A. JOHN BAILER ¹, ²
SUZANNE R. KUNKEL ¹, ³

1. SCRIPPS GERONTOLOGY CENTER
2. DEPARTMENT OF STATISTICS
3. DEPARTMENT OF SOCIOLOGY & GERONTOLOGY

MIAJNI UNIVERSITY
OUTLINE

1. Gapminder.org
2. Data Visualization Practice in Gerontology
3. General Guideline for “Good” Data Visualization Practice
4. R graphics demonstration
5. Q & A
How would you present the following data?

• Life expectancy (years) and income (GDP per capita) by nations
• About 200 years
• About 250 countries and country-equivalent areas

Example:
Gampminder.org
Gapminder Desktop
GAPMINDER.ORG

• Did you look at the interactive bubble chart?
• Did you like it?
• What did you like about the Gapminder World example?
• Should gerontologists use “good” data graphics? Research? Practice? Education?
WHAT DO GERONTOLOGISTS DO?

“A descriptive review of graphics in The Gerontologist published first 10 years of 21st century: recent practice, suggested guideline and future direction”

Takashi Yamashita, A. John Bailer & Suzanne R. Kunkel
INTRODUCTION

• Larger and more complex data <- population aging
• Volume & complexity (e.g., demographic characteristics, SES)
• Health status of older adults by sex, education, race, regions….
• Better understanding of data <- good presentation of data <- well-designed data graphics

Research questions
1. What kind of graphics do gerontologists use?
2. What should we pay attention to when designing data graphics (illustrative examples)?
METHODS

  • All articles (n = 1,189)
  • All graphics in four types of articles including original research, brief report, the forum and practice concepts (n = 863)

• Exhaustive search of graphics
• Classification (Robbins, 2005)
• Illustrative examples & discussion (Cleveland, 1985)
GRAPHICS CLASSIFICATION

1. Bar chart
2. Box plot
3. Dot plot
4. Histogram
5. Line graph
6. Map
7. Mosaic plot
8. Pie chart
9. Stacked bar chart
10. Grouped bar chart
11. Strip chart
12. Scatter plot
13. Scatter plot matrix
14. Flow chart
15. Others (e.g., conceptual model)

Robbins (2005)
## RESULTS (N = 863)

<table>
<thead>
<tr>
<th>N of graphic per article</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>548 (63.5%)</td>
</tr>
<tr>
<td>1</td>
<td>165 (19.1%)</td>
</tr>
<tr>
<td>2</td>
<td>81 (9.4%)</td>
</tr>
<tr>
<td>3</td>
<td>35 (4.1%)</td>
</tr>
<tr>
<td>4</td>
<td>18 (2.1%)</td>
</tr>
<tr>
<td>5</td>
<td>11 (1.3%)</td>
</tr>
<tr>
<td>6</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>7</td>
<td>1 (&lt; 1.0%)</td>
</tr>
<tr>
<td>8</td>
<td>3 (&lt; 1.0%)</td>
</tr>
<tr>
<td>9</td>
<td>1 (&lt; 1.0%)</td>
</tr>
</tbody>
</table>

**Total n of graphics** 599
RESULTS (N = 863)

64% of articles used NO graphics

30% of articles used 1, 2 or 3 graphics
Gerontologists' Choice #3

Grouped bar chart

Herd P. The Gerontologist 2005;45:12-25

Levy-Storms L et al. The Gerontologist 2007;47:14-20

![Grouped bar chart](image-url)
GERONTOLOGISTS’ CHOICE

#2

Line Graph (interaction plot; time series; model fit)


Hernandez M The Gerontologist 2007;47:118-124

GERONTOLOGISTS’ CHOICE

#1 Others (Conceptual Model, Flow Chart, SEM diagrams)


Tang F et al. The Gerontologist 2010;50:603-612
### GERONTOLOGISTS’ CHOICE

<table>
<thead>
<tr>
<th>Type of graphics</th>
<th>Category</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line graph</td>
<td></td>
<td>204 (34.1%)</td>
</tr>
<tr>
<td>Grouped bar chart</td>
<td></td>
<td>82 (13.7%)</td>
</tr>
<tr>
<td>Bar chart</td>
<td></td>
<td>33 (5.5%)</td>
</tr>
<tr>
<td>Pie chart</td>
<td></td>
<td>11 (1.8%)</td>
</tr>
<tr>
<td>Stacked bar chart</td>
<td></td>
<td>11 (1.8%)</td>
</tr>
<tr>
<td>Dot plot</td>
<td></td>
<td>6 (1.0%)</td>
</tr>
<tr>
<td>Histogram</td>
<td></td>
<td>6 (1.0%)</td>
</tr>
<tr>
<td>Scatter plot</td>
<td></td>
<td>6 (1.0%)</td>
</tr>
<tr>
<td>Map</td>
<td></td>
<td>3 (0.5%)</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>237 (39.6%)</td>
</tr>
</tbody>
</table>
GERONTOLOGISTS' CHOICE

Bar Chart
Dot Plot
Grouped Bar Chart
Histogram
Line Chart
Map
Others
Pie Chart
Scatter Plot
Stacked Bar Chart

Number of Graphics
GERONTOLOGISTS’ CHOICE

- **Line Chart**: 204 (34%)
- **Grouped Bar Chart**: 82 (14%)
- **Bar Chart**:
- **Pie Chart**:
- **Stacked Bar Chart**:
- **Dot Plot**:
- **Histogram**:
- **Scatter Plot**:
- **Map**:
- **Others**: 237 (40%)

Number of Graphics
SUMMARY

- 64% (548 / 863) of TG articles used NO graphics
- 89% of articles used tables (2,746 tables!)
- 599 graphics in 315 TG articles
- Gerontologists’ choice of graphics
  1. Others (e.g., conceptual model, flow chart)
  2. Line graph
  3. Grouped bar chart
- Gerontologists focus on theoretical/conceptual framework
- Data graphics are not the conclusion but resources for better presentation/communication and decision making
Better

1. Position along a common scale

2. Position along nonaligned scales

3. Length

4. Angle-slope

5. Area

6. Volume

7. Color

Worse
**EXAMPLE DATA: HOW WOULD YOU PRESENT THE DATA?**

<table>
<thead>
<tr>
<th>Age group</th>
<th>Little disability (%)</th>
<th>Moderate disability (%)</th>
<th>Severe disability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65-69</td>
<td>77</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>70-74</td>
<td>70</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>75-79</td>
<td>65</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>80-84</td>
<td>61</td>
<td>24</td>
<td>15</td>
</tr>
<tr>
<td>85-89</td>
<td>46</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>90-94</td>
<td>34</td>
<td>28</td>
<td>38</td>
</tr>
<tr>
<td>95+</td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

Estimated distribution of disability status in Ohio’s older populations by age group, 1995
Estimated distribution of disability status in Ohio’s older populations by age group, 1995
Estimated Distribution of Disability Status in Ohio's Older population by age, 1995

- Little or No Disability is DECREASING with age
- Severe Disability is INCREASING with age

Note: Moderate disability increases slightly with age.
Ohio Long-Term Care Research Project Report. Scripps Gerontology Center, Miami University, Oxford, OH 45056
STATISTICAL GRAPHICS – COMMON OPTIONS AND CURRENT TOOLS
CHARTS PICKED FROM MENU
NEW GRAPHICS IDEAS –
“GRAMMAR OF GRAPHICS”

Wickham (2009) – ggplot2

Statistical graphic = mapping of DATA to AESTHETIC attributes of GEOMETRIC objects

Built by layers for objects rendered on plot
Data + aesthetic mapping

“geoms” = geometric objects what you see on a plot – e.g. points, lines, polygons

“stats” – stat transf. that summarizes data (e.g. binning for histogram)

“scales” – map of data space value to aesthetic space value

“coord” – coordinate system (e.g. Cartesian, polar, map projections)

“faceting” – breaking data into subsets
WHAT IS A SCATTERPLOT?

Data: (x,y) pairs

Geom: point

Scale: map (x,y) data value units to graphic page physical units (generates axis and legend)

Coord: Cartesian system
AN EXAMPLE OF GROWTH

str(growth.df)
'data.frame': 100 obs. of 7 variables:
$ obs     : num 1 2 3 4 5 6 7 8 9 10 ...
$ response: num 111 116 122 126 130 ... # height of kid
$ child    : num 1 1 1 1 1 2 2 2 2 2 ... # 5 measurements on each kid
$ age      : num 6 7 8 9 10 6 7 8 9 10 ... # at ages 6, 7, 8, 9, 10
$ group    : num 1 1 1 1 1 1 1 1 1 1 ...
$ Fchild   : Factor w/ 20 levels "1","2","3","4",...: 1 1 1 1 1 2 2 2 2 2 ...
$ Fgroup   : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
# group = height of mom
SCATTERPLOT

```r
scat.plot <- qplot(x=age,y=response,data=growth.df)
```

```r
names(scat.plot)
```

```
[1] "data"      "layers"    "scales"    "mapping"   "options"
[6] "coordinates" "facet"     "plot_env"
```

```r
scat.plot$layers:
geom_point:
stat_identity:
position_identity: (width = NULL, height = NULL)
```
qplot(x=age, y=response, group=child, data=growth.df) + geom_point(aes(shape=child, colour=Fgroup)) + geom_line(aes(colour=Fgroup, linetype=group))
p <- ggplot(aes(x=age,y=response),data=growth.df)
p1 <- p+geom_point(aes(colour=Fgroup,shape=child)) + 
  facet_wrap(~Fchild) + 
  geom_smooth(aes(colour=Fgroup),method="lm",se=FALSE)
p + geom_point(aes(colour=Fgroup, shape=child)) + facet_wrap(~Fchild) + geom_smooth(aes(colour=Fgroup), method="lm", se=FALSE) + theme_bw()
WHAT HAS CHANGED?

names(p1)
[1] "data" "layers" "scales" "mapping" "options"
[6] "coordinates" "facet" "plot_env"

p1$layers
[[1]]
  mapping: colour = Fgroup, shape = child
  geom_point: na.rm = FALSE
  stat_identity:
  position_identity: (width = NULL, height = NULL)

[[2]]
  mapping: colour = Fgroup
  geom_smooth:
  stat_smooth: method = lm, se = FALSE
  position_identity: (width = NULL, height = NULL)
RESOURCES

http://had.co.nz/ggplot2/
http://learnr.wordpress.com/tag/ggplot2/
http://www.slideshare.net/izahn/rgraphics-12040991
FUTURE? (AND PRESENT)

Easier interaction with plots
Dynamic displays of plots (ggobi)
Browser-based displays
Text visualization
Easier and more available display tools
RESOURCES (BEYOND TUFTE) ...


ASA Section on Statistical Computing and Statistical Graphics (http://stat-computing.org/newsletter )

Sparklines – small intense, simple, word-sized graphic with typographic resolution (Tufte 2004)

QUESTIONS/COMMENTS/

Thank you for your attention and interest.
Hard questions should be addressed to Taka and Suzanne.
Easy questions should be addressed to John.