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

MIAMI UNIVERSITY

SCRIPPS GERONTOLOGY CENTER
AN OHIO CENTER OF EXCELLENCE

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**

GAPMINDER.ORG

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**

<u>Research questions</u>

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

METHODS

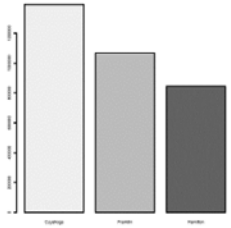
- **Data: all articles published in *The Gerontologist (TG)* between 2001 and 2010.**
  - **All articles (n = 1,189)**
  - **All graphics in four types of articles including original research, brief report, the forum and practice concepts (n = 863)**
- **Exhaustive search of graphics**
- **Classification (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)

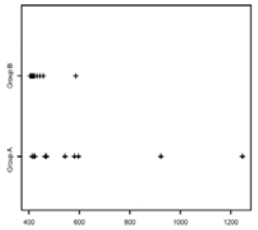
Robbins (2005)

Bar Chart

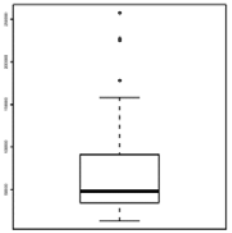


Map

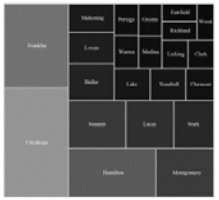
Strip Chart



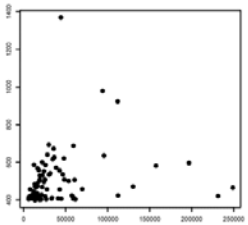
Box Plot



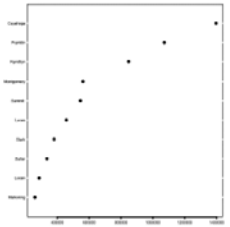
Mosaic Plot



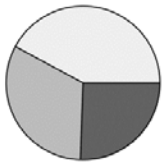
Scatter Plot



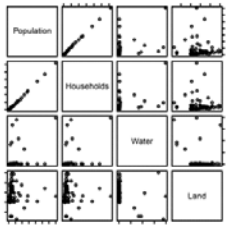
Dot Plot



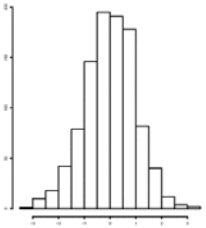
Pie Chart



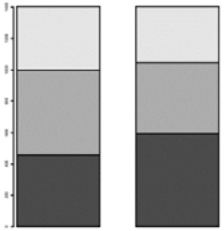
Scatter Plot Matrix



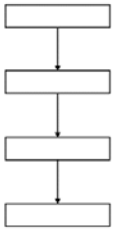
Histogram



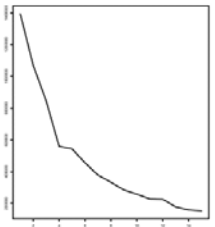
Stacked Bar Chart



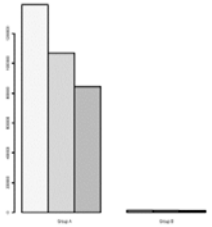
Others (Flowchart)



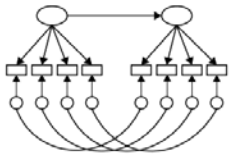
Line Graph



Grouped Bar Chart



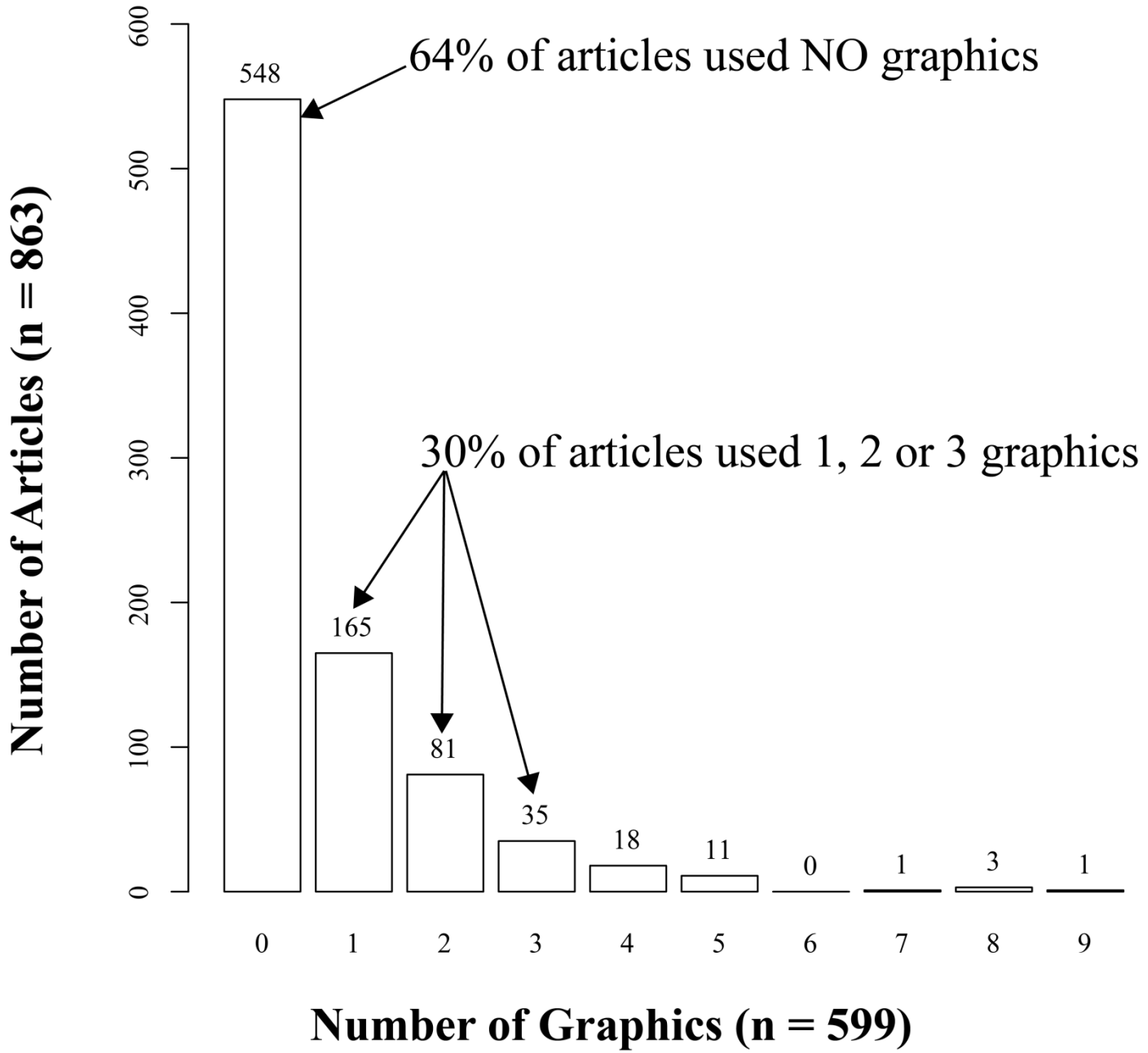
Others
(Conceptual Model)



RESULTS (N = 863)

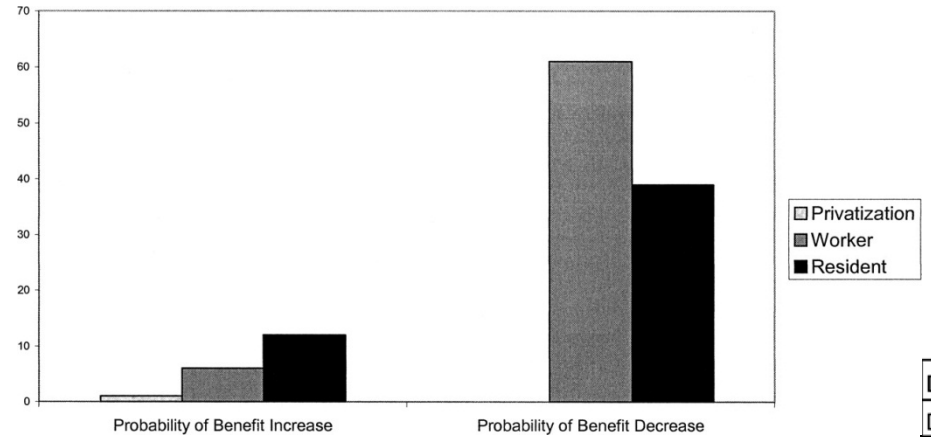
N of graphic per article	N (%)
0	548 (63.5%)
1	165 (19.1%)
2	81 (9.4%)
3	35 (4.1%)
4	18 (2.1%)
5	11 (1.3%)
6	0 (0%)
7	1 (< 1.0%)
8	3 (< 1.0%)
9	1 (< 1.0%)
Total n of graphics	599

RESULTS (N = 863)

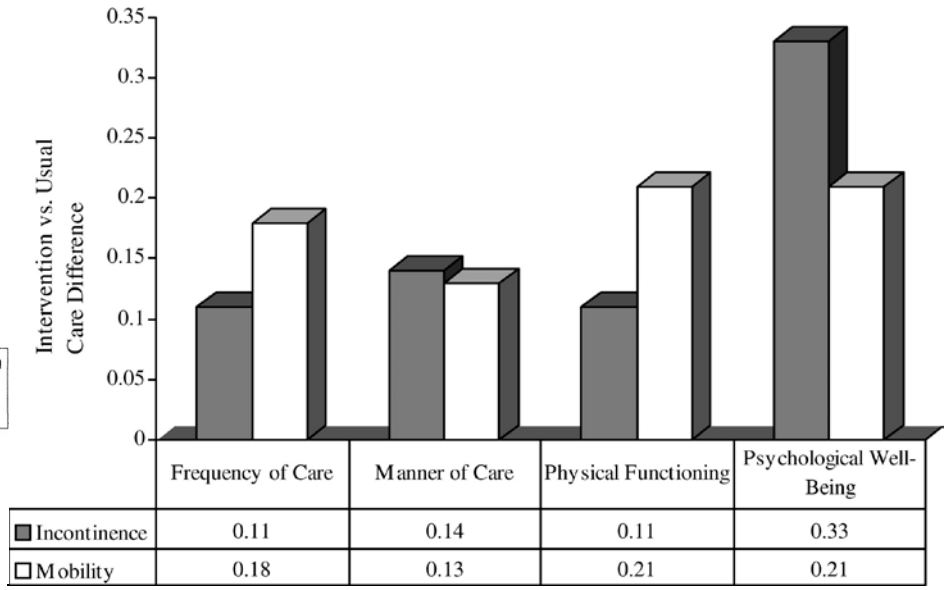


GERONTOLOGISTS' CHOICE #3

Grouped bar chart



Herd P The Gerontologist 2005;45:12-25

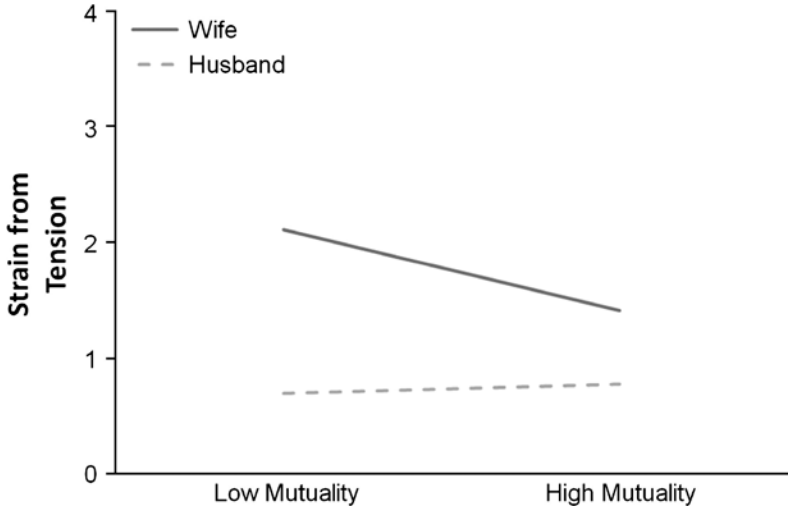


	Frequency of Care	Manner of Care	Physical Functioning	Psychological Well-Being
Incontinence	0.11	0.14	0.11	0.33
Mobility	0.18	0.13	0.21	0.21

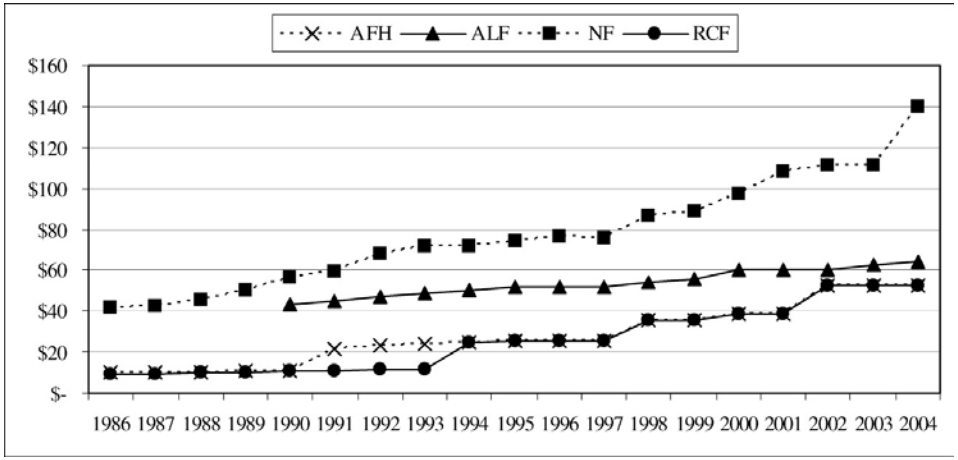
Levy-Storms L et al. The Gerontologist 2007;47:14-20

GERONTOLOGISTS' CHOICE #2

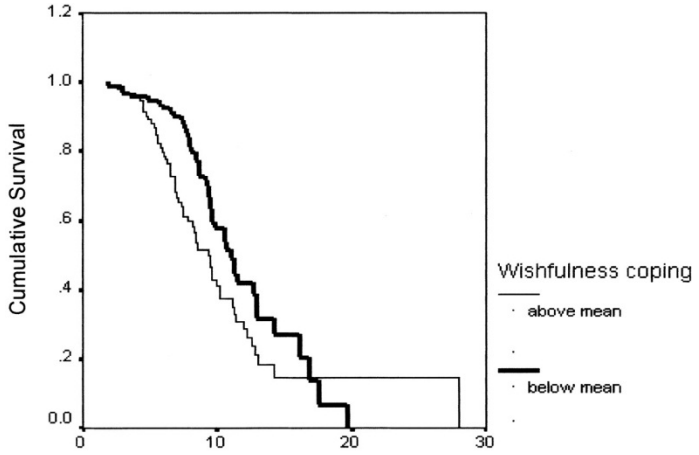
Line Graph (interaction plot; time series; model fit)



Lyons K S et al. The Gerontologist 2009;49:378-387



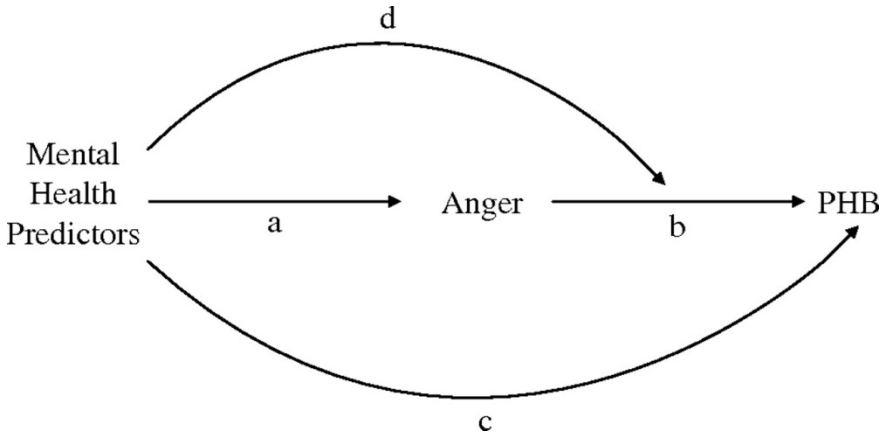
Hernandez M The Gerontologist 2007;47:118-124



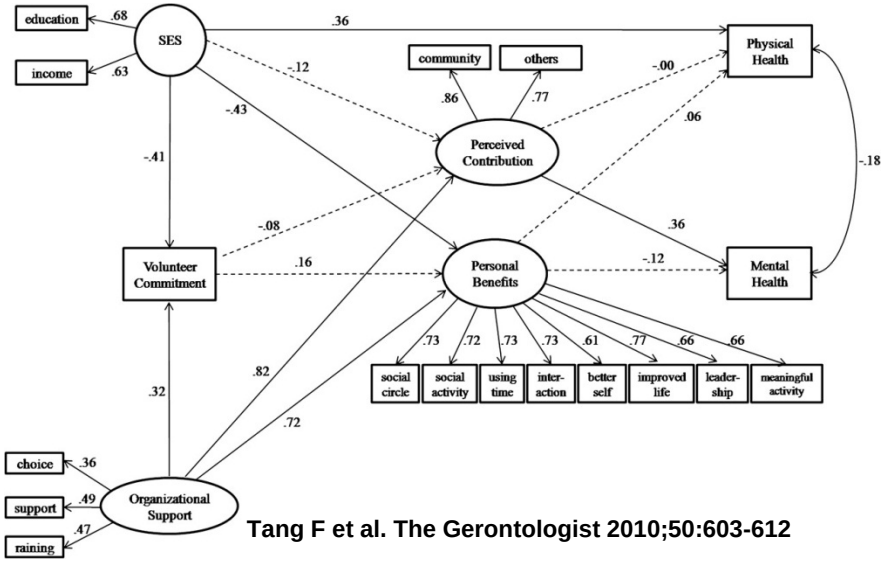
McClendon M J et al. The Gerontologist 2004;44:508-519

GERONTOLOGISTS' CHOICE #1

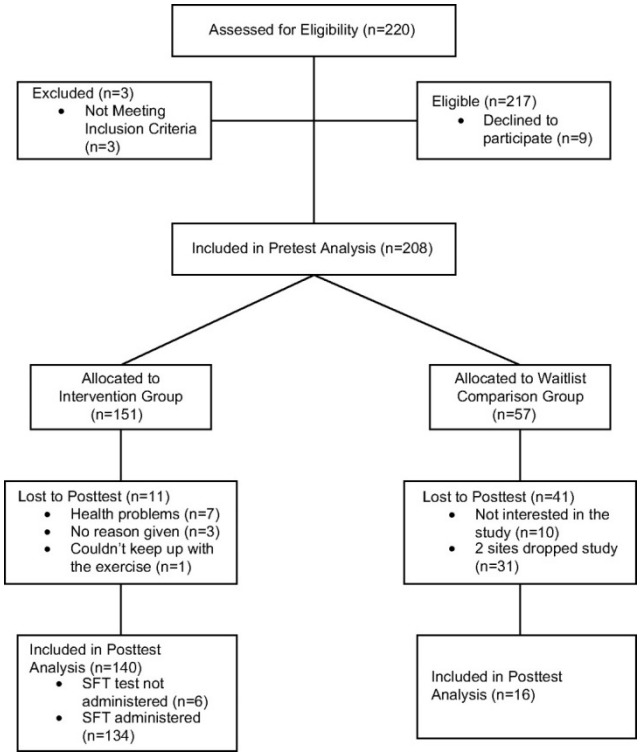
Others (Conceptual Model, Flow Chart, SEM diagrams)



MacNeil G et al. The Gerontologist 2009;50:76-86



Tang F et al. The Gerontologist 2010;50:603-612

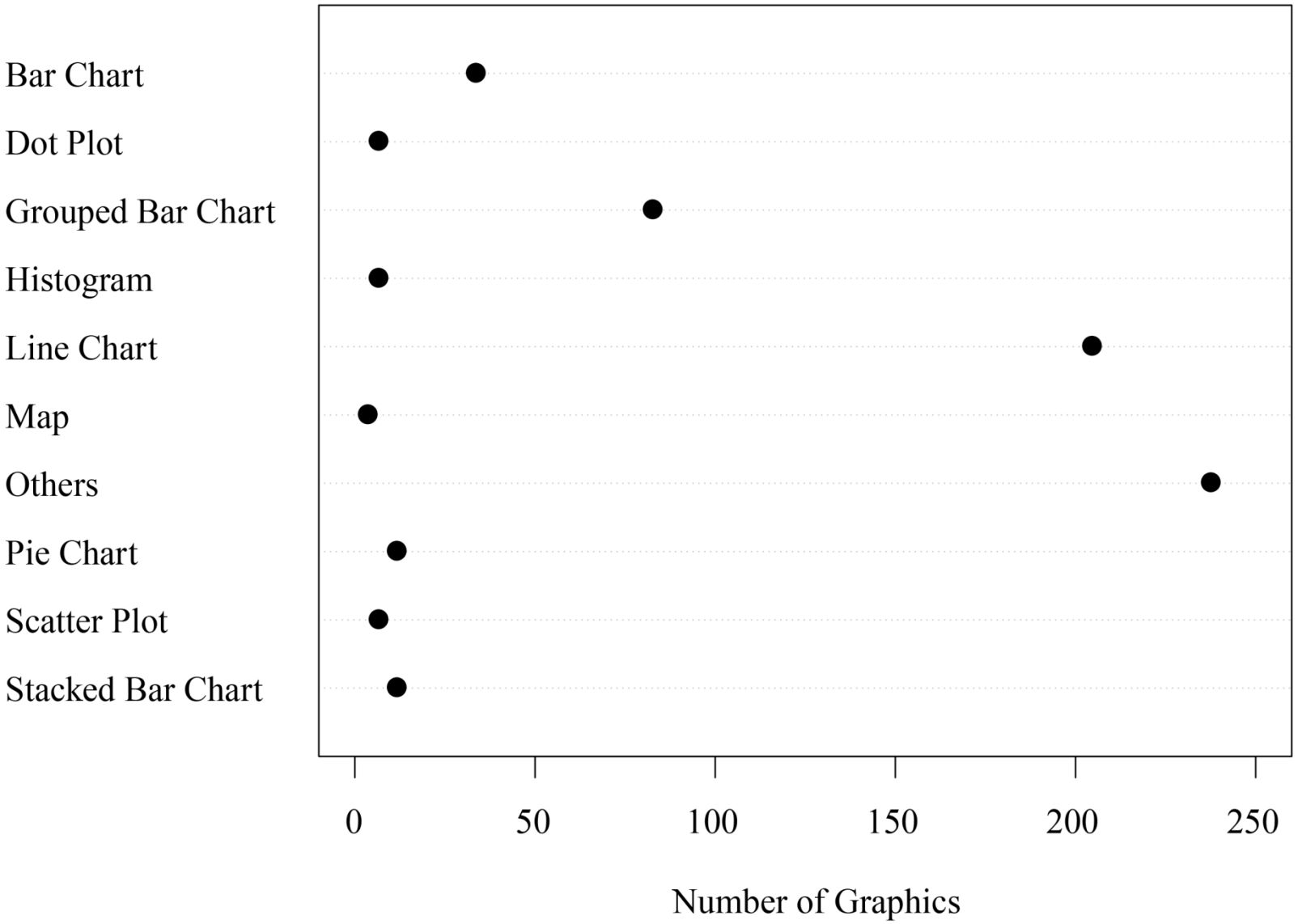


Yan T et al. The Gerontologist 2009;49:847-855

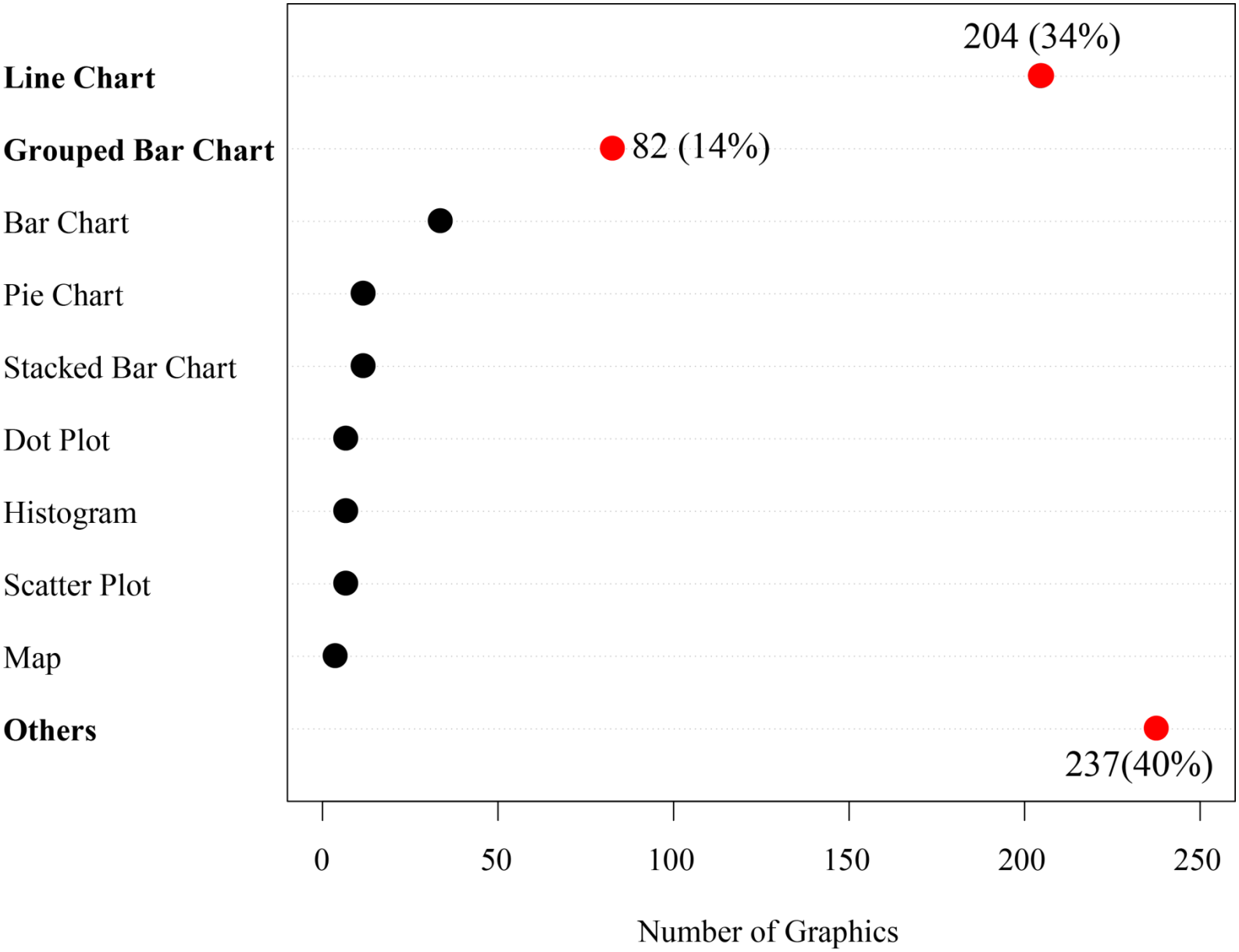
GERONTOLOGISTS' CHOICE

Type of graphics	Category
	n (%)
Line graph	204 (34.1%)
Grouped bar chart	82 (13.7%)
Bar chart	33 (5.5%)
Pie chart	11 (1.8%)
Stacked bar chart	11 (1.8%)
Dot plot	6 (1.0%)
Histogram	6 (1.0%)
Scatter plot	6 (1.0%)
Map	3 (0.5%)
Others	237 (39.6%)

GERONTOLOGISTS' CHOICE



GERONTOLOGISTS' CHOICE



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**

CLEVELAND VISUAL ELEMENTS HIERARCHY

Better

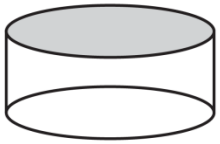
1. Position along a common scale

2. Position along nonaligned scales

3. Length

4. Angle-slope

5. Area



6. Volume

Worse

7. Color

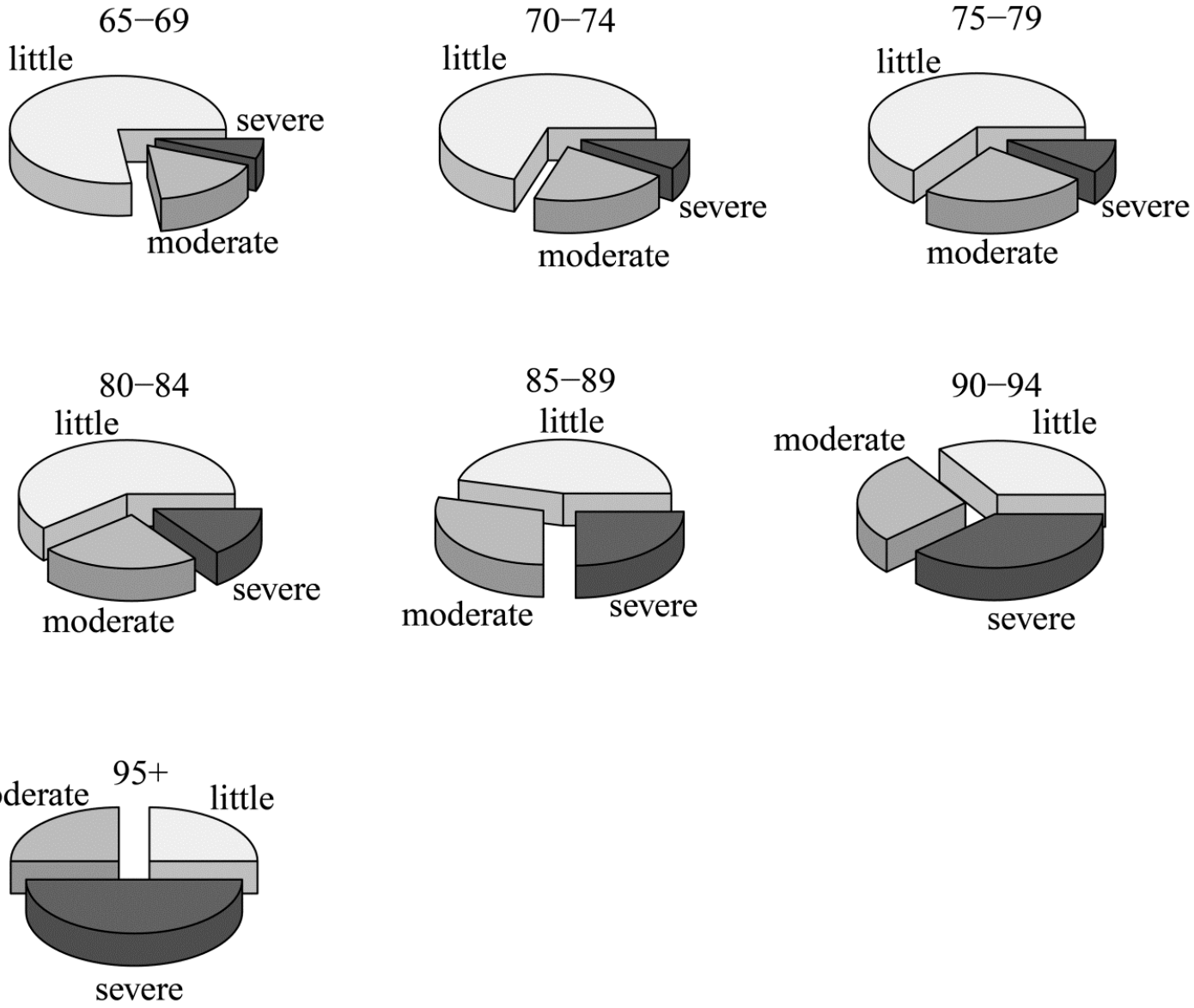
EXAMPLE DATA: HOW WOULD YOU PRESENT THE DATA?

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

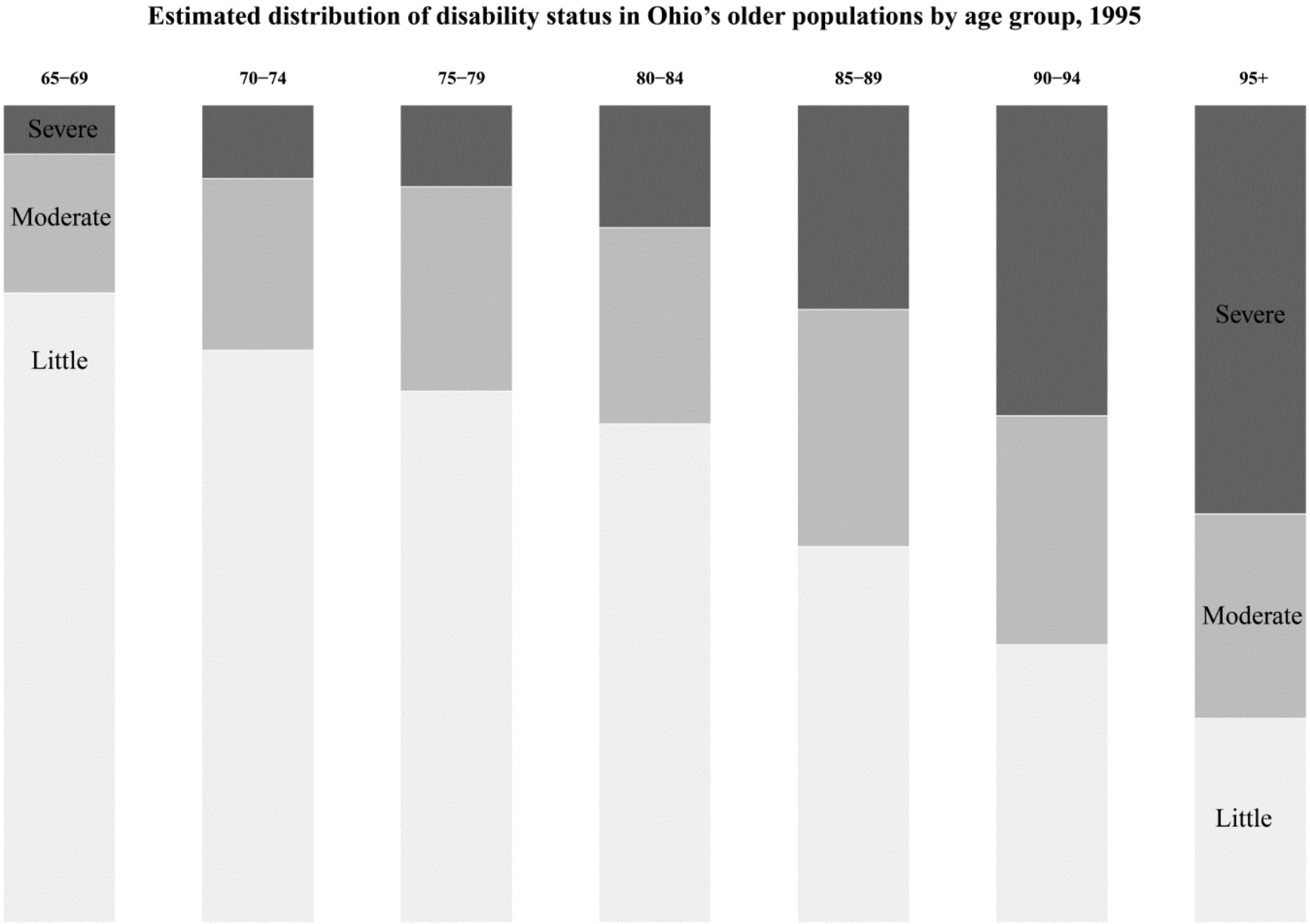
Data Source: Mehdizadeh et al. (2001). Projection of Ohio's Older Population: 2015-2050 Ohio Long-Term Care Research Project Report. Scripps Gerontology Center, Miami University, OH 45056

EXAMPLE #1

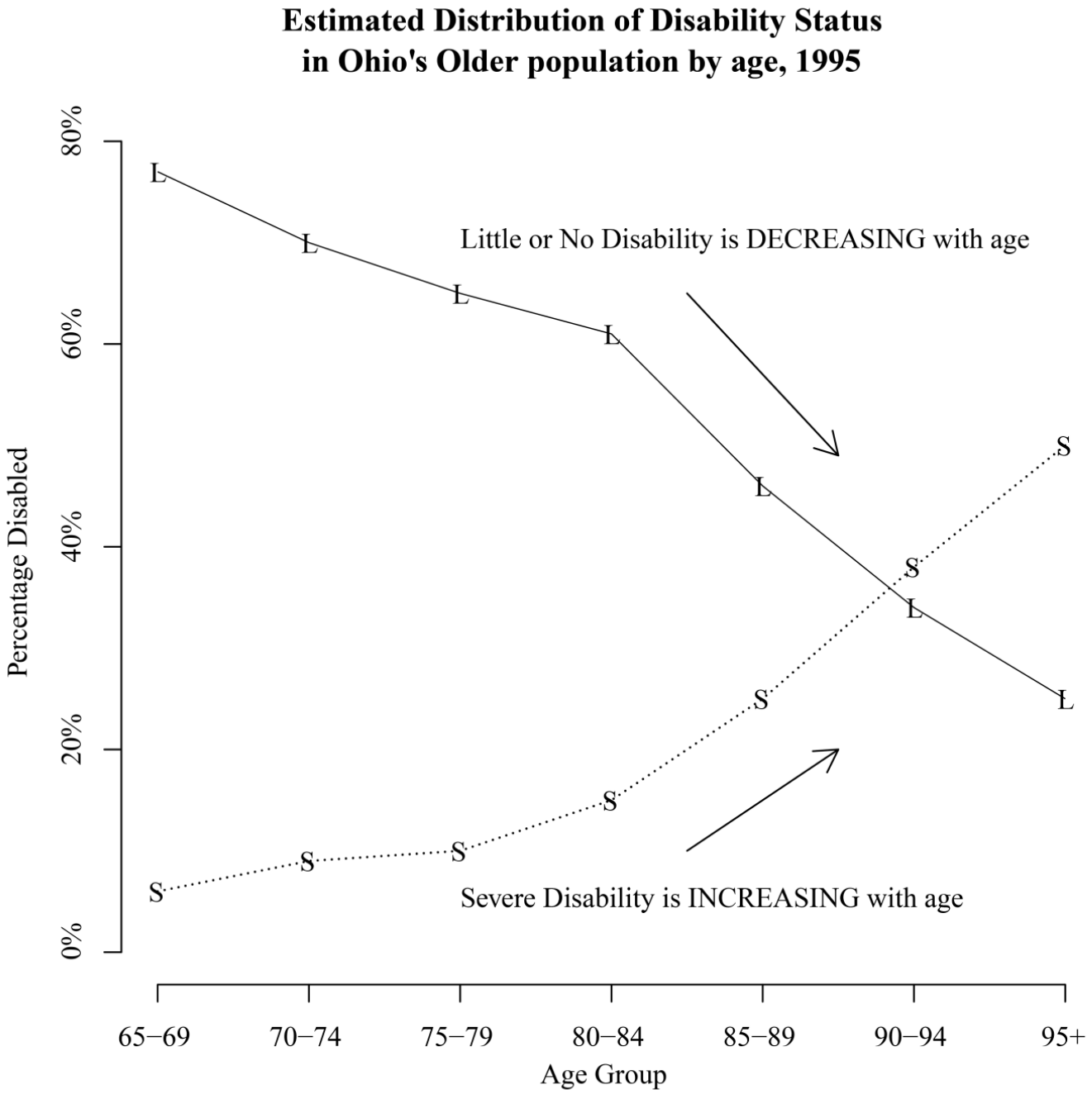
Estimated distribution of disability status in Ohio's older populations by age group, 1995



EXAMPLE #2


Estimated distribution of disability status in Ohio's older populations by age group, 1995
65–69
70–74
75–79
80–84
85–89
90–94
95+
Severe
Moderate
Little
Severe
Moderate
Little

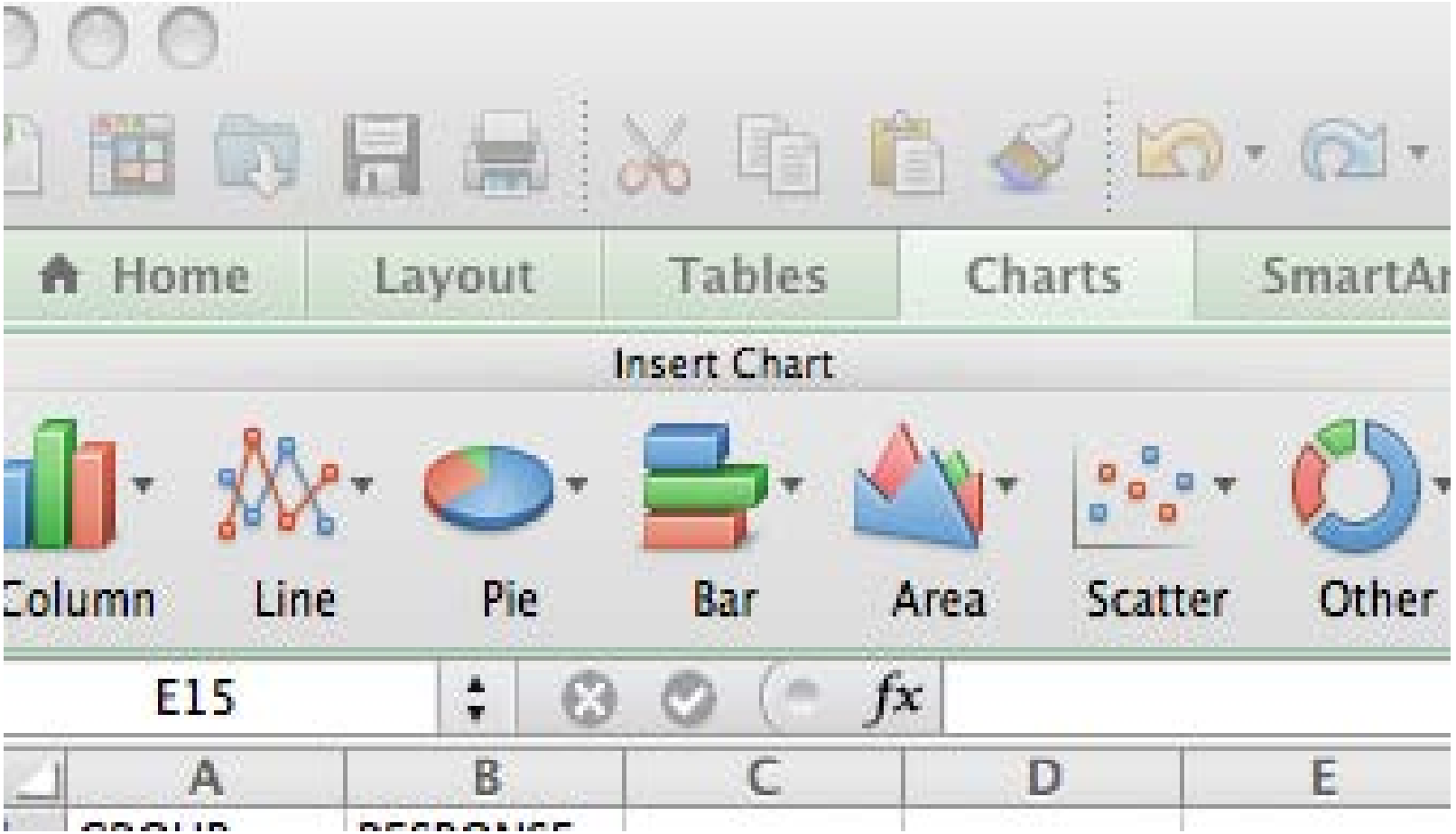
EXAMPLE #3



Note: Moderate disability increases slightly with age.
Source: Mehdizadeh et al. (2001) Projections of Ohio's Older Disabled Population: 2015 to 2050
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"

Wilkinson (2003)

Wickham (2009) – ggplot2

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

Built by layers for objects rendered on plot

PIECES …

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)

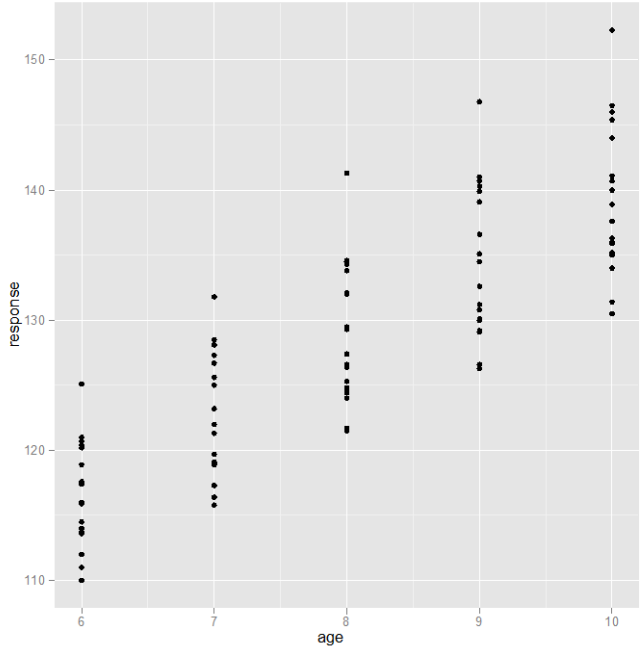
Coord: Cartesian system

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
```

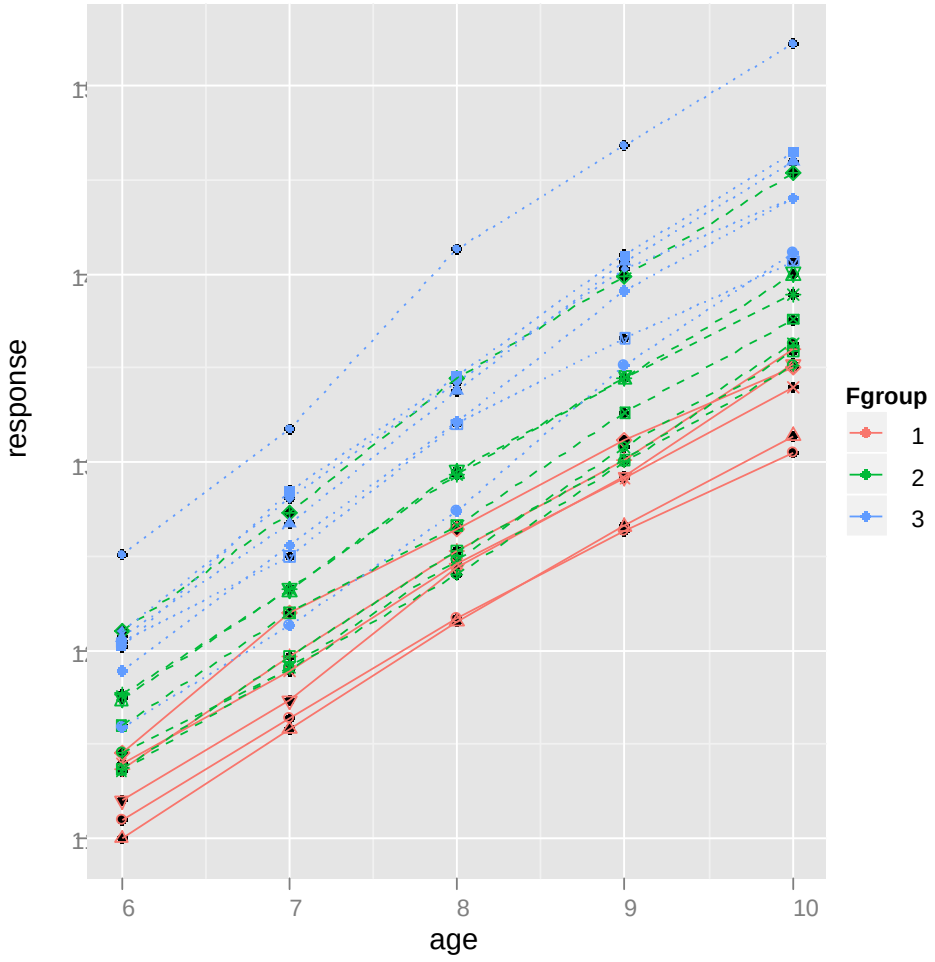
SCATTERPLOT

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

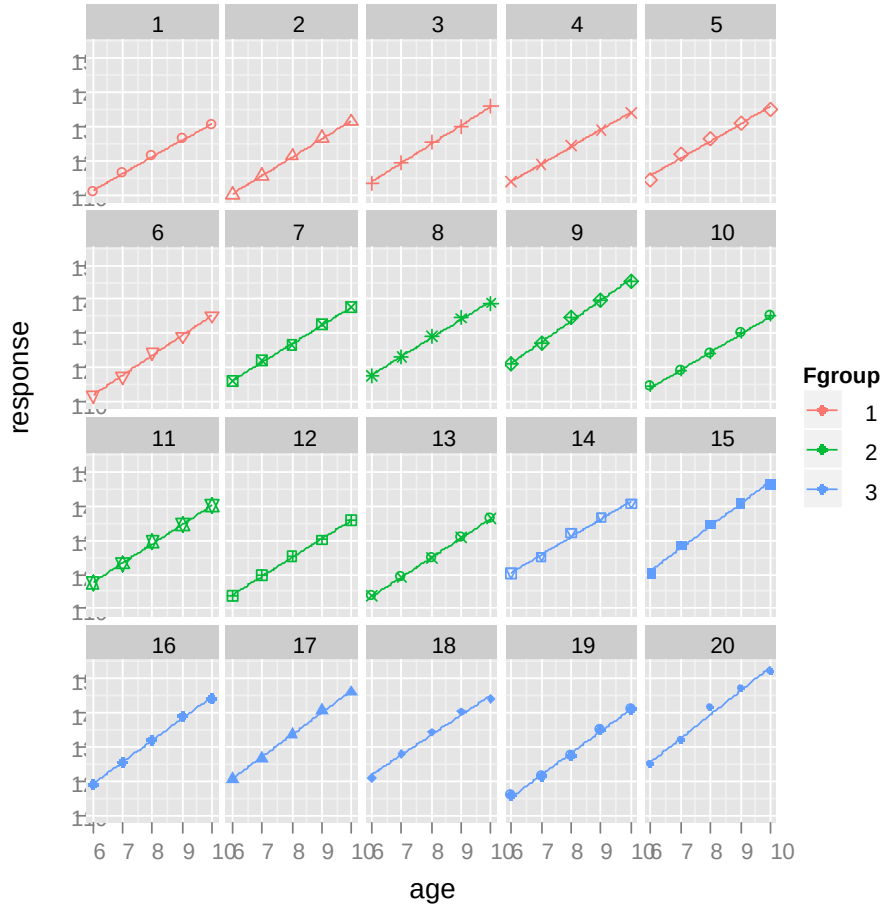



```
names(scat.plot)
[1] "data"        "layers"      "scales"      "mapping"     "options"
[6] "coordinates" "facet"       "plot_env"
scat.plot$layers:
geom_point:
stat_identity:
position_identity: (width = NULL, height = NULL)
```

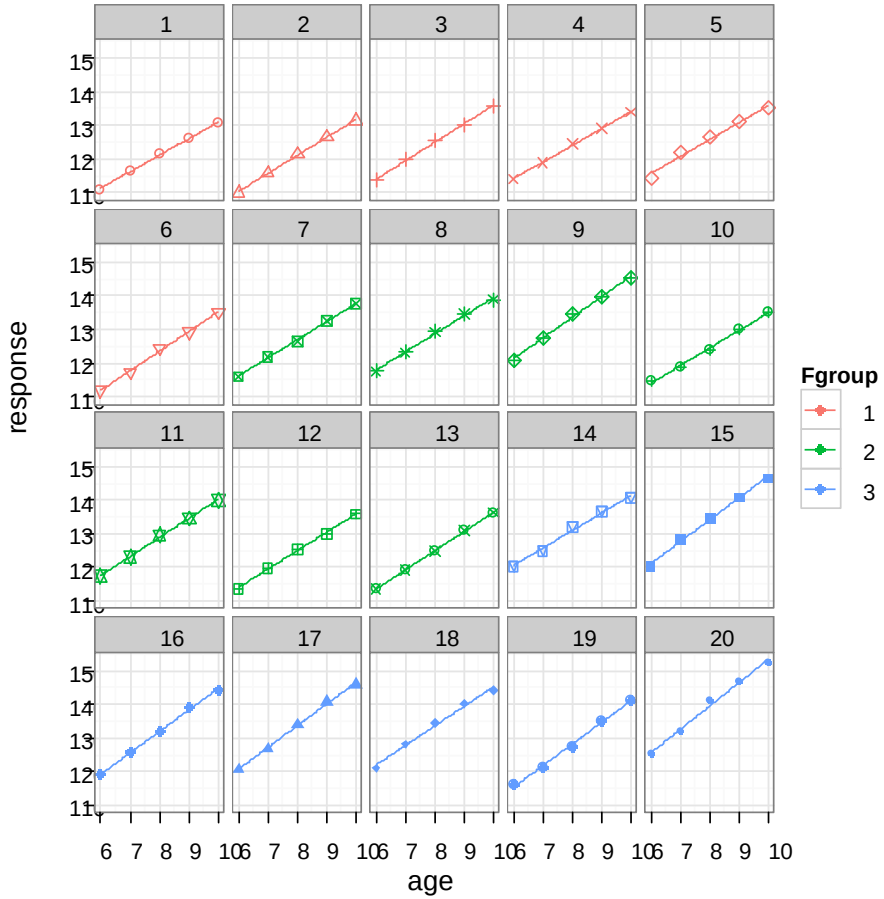
```
qplot(x=age,y=response,group=child,data=growth.df) +
    geom_point(aes(shape=child,colour=Fgroup)) +
    geom_line(aes(colour=Fgroup,linetype=group))
```



```
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) …

The Golden Age of Statistical Graphics -- Michael Friendly http://arxiv.org/pdf/0906.3979.pdf

Wainer H, Velleman PF (2001) Statistical graphics: mapping the pathways of science. ANNUAL REVIEW OF PSYCHOLOGY 52: 305-335.

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)

http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR

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.