431 Class 11

Thomas E. Love

2018-10-02

Today's Agenda

1. The Western Collaborative Group Study
   - Loose End 1: Adding Notches to Boxplots
   - Loose End 2: Standard Error of the Mean
   - Loose End 3: Identification of Missingness
   - Studying Associations: Correlation Matrix
   - Studying Associations: Scatterplot Matrix
2. Some Thoughts on Building Tables Well
3. A little more on Leek Chapters 3, 4, 12

Today's R Starting Point

```
library(GGally); library(tidyverse)
```

Western Collaborative Group Study (wcgs)

See Notes, Chapter 13.

- Full data set has 3,154 observations on 22 variables.

```
wcgs.full <- read.csv("wcgs.csv") %>% tbl_df()

set.seed(4312018)

wcgs1 <- wcgs.full %>%
  mutate(bmi2 = round(bmi, 2), subj = as.character(id)) %>%
  select(subj, age, chol, bmi2, smoke, ncigs,
         behpat, chd69) %>%
  sample_n(size = 500)
```

```
dim(wcgs1)
```

```
[1] 500   8
```

What's in wcgs1?

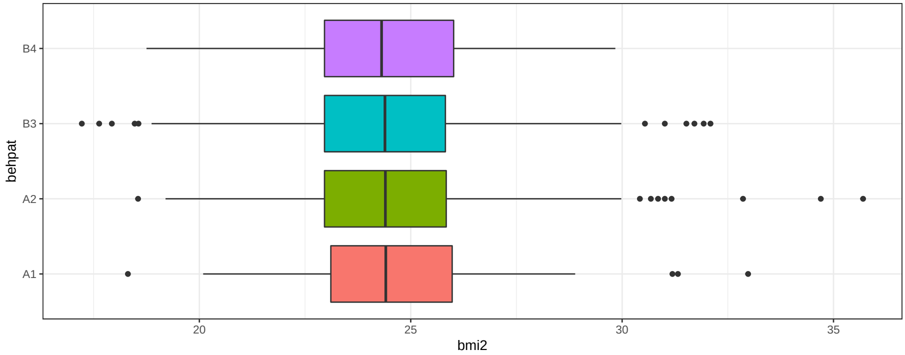
- subj = subject identification code
- age (in years)
- chol is total cholesterol in mg/dl
- bmi2 is body-mass index, rounded to two decimal places
- smoke is Yes if cigarette smoker, No if not
- ncigs is # of cigarettes smoked/day
- behpat is behavioral pattern (A1, A2, B3 or B4)
- chd69 is whether subject had a CHD event (Yes/No)

```
slice(wcgs1, 105:107)
```

```
# A tibble: 3 x 8
  subj    age  chol  bmi2 smoke ncigs behpat chd69
  <chr> <int> <int> <dbl> <fct> <int> <fct>  <fct>
1 11513    39   168  24.4 No        0 B3     No
2 21242    39   322  25.1 Yes      25 A2     No
3 3498     56   237  25.5 No        0 A1     Yes
```

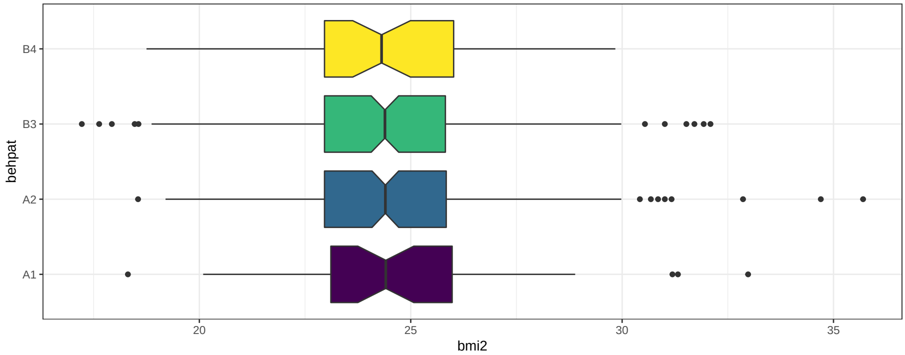
A Loose End: Comparison Boxplots

```
ggplot(wcgs1, aes(x = behpat, y = bmi2, fill = behpat)) +
  geom_boxplot() +
  coord_flip() + guides(fill = FALSE) + theme_bw()
```



A Loose End: Adding Notches to Comparison Boxplots

```
ggplot(wcgs1, aes(x = behpat, y = bmi2, fill = behpat)) +
  geom_boxplot(notch = TRUE) +
  scale_fill_viridis_d() +
  coord_flip() + guides(fill = FALSE) + theme_bw()
```



A Loose End: Standard Error of the Sample Mean

```r
mosaic::favstats(chol ~ chd69, data = wcgs1)
```

```
  chd69 min  Q1 median  Q3 max    mean       sd   n
1    No 129 194    220 256 400 225.529 44.32924 448
2   Yes 171 201    235 259 318 234.160 36.34885  50
  missing
1       2
2       0
```

The standard error for the `No` group is

$$SE(\bar{x}) = \frac{SD}{\sqrt{n}} = \frac{44.32924}{\sqrt{448}} = \frac{44.32924}{21.16601} = 2.09$$

Comparing the Standard Errors using the tidyverse

```
wcgs1 %>%
  filter(complete.cases(chd69, chol)) %>%
  group_by(chd69) %>%
  summarize(n = n(), mean(chol), sd(chol),
            "se(chol)" = sd(chol)/(sqrt(n()))) %>%
  knitr::kable()
```

chd69	n	mean(chol)	sd(chol)	se(chol)
No	448	225.529	44.32924	2.094360
Yes	50	234.160	36.34885	5.140504

Missing Data?

```
wcgs1 %>%
  summarize_all(funs(sum(is.na(.))))
```

```
# A tibble: 1 x 8
   subj   age  chol  bmi2 smoke ncigs behpat chd69
  <int> <int> <int> <int> <int> <int>  <int> <int>
1     0     0     2     0     0     0      0     0
```

Or, use `summary`, or `mosaic::favstats` or other approaches.

Could use the `map` approach from the `purrr` package:

```
map(wcgs1, ~sum(is.na(.)))
```

```
$subj
[1] 0


$age
[1] 0
```

Which rows have missing data?

```
wcgs1 %>%
  filter(!complete.cases(.))
```

```
# A tibble: 2 x 8
  subj    age  chol  bmi2 smoke ncigs behpat chd69
  <chr> <int> <int> <dbl> <fct> <int> <fct>  <fct>
1 13294    56    NA  25.6 Yes      10 A1     No
2 12239    45    NA  26.3 No        0 B4     No
```

New Tools: The Correlation Matrix and the Scatterplot Matrix

A Correlation Matrix for the Quantitative Variables

```
wcgs1 %>%
  select(chol, age, bmi2, ncigs) %>%
  cor() %>%
  round(., 3) %>%
  knitr::kable()
```

	chol	age	bmi2	ncigs
chol	1	NA	NA	NA
age	NA	1.000	-0.065	-0.045
bmi2	NA	-0.065	1.000	-0.114
ncigs	NA	-0.045	-0.114	1.000

A Correlation Matrix for the Quantitative Variables

Accounting for missingness by dropping incomplete cases...

```
wcgs1 %>%
  select(chol, age, bmi2, ncigs) %>%
  filter(complete.cases(.)) %>%
  cor() %>%
  round(., 3) %>%
  knitr::kable()
```

	chol	age	bmi2	ncigs
chol	1.000	0.083	0.079	0.137
age	0.083	1.000	-0.066	-0.046
bmi2	0.079	-0.066	1.000	-0.113
ncigs	0.137	-0.046	-0.113	1.000

All these correlations are based on 498 observations, rather than 500.

A Correlation Matrix for the Quantitative Variables

What if we want the `chol`-based correlations to use 498, but the rest to use all of the data (500 observations)?

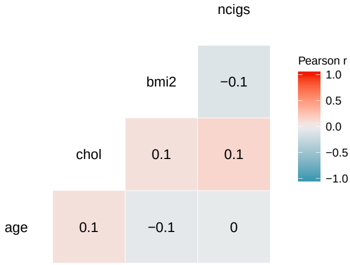
```
wcgs1 %>%
  select(chol, age, bmi2, ncigs) %>%
  cor(., use = "pairwise.complete.obs") %>%
  round(., 3) %>%
  knitr::kable()
```

	chol	age	bmi2	ncigs
chol	1.000	0.083	0.079	0.137
age	0.083	1.000	-0.065	-0.045
bmi2	0.079	-0.065	1.000	-0.114
ncigs	0.137	-0.045	-0.114	1.000

Using `ggcorr` from `GGally` for a Correlation Matrix

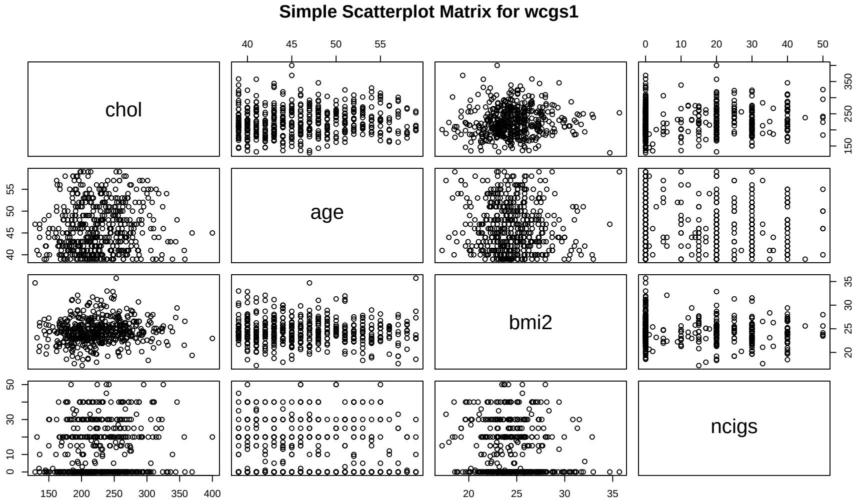
```r
ggcorr(wcgs1, name = "Pearson r", label = TRUE)
```

```
Warning in ggcorr(wcgs1, name = "Pearson r", label =
TRUE): data in column(s) 'subj', 'smoke', 'behpat',
'chd69' are not numeric and were ignored
```



A Scatterplot Matrix for the Numeric Variables

```r
pairs(~ chol + age + bmi2 + ncigs, data = wcgs1,
      main = "Simple Scatterplot Matrix for wcgs1")
```

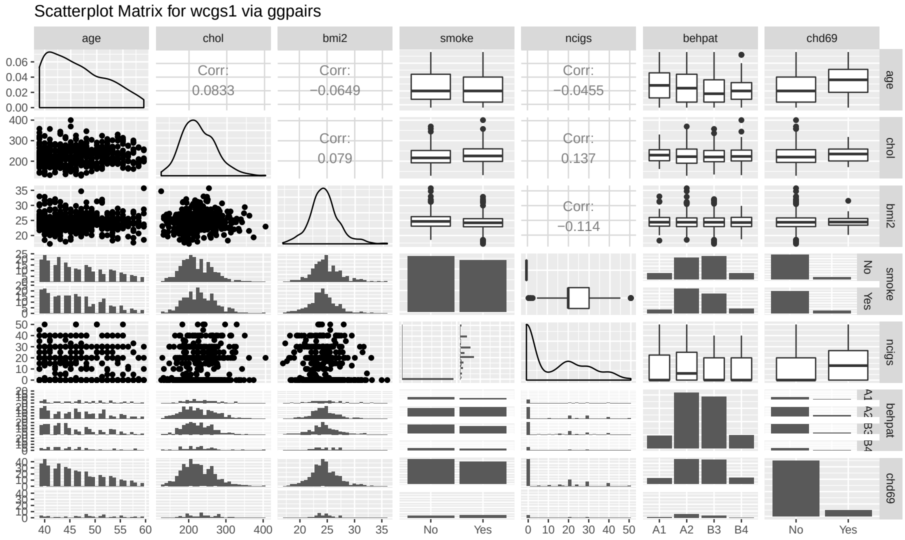


Scatterplot Matrix via ggpairs in GGally (Code)

```
ggpairs(wcgs1 %>% select(-subj),
        title = "Scatterplot Matrix for wcgs1 via ggpairs")
```

- In practice, I run this with `warning = FALSE` and `message = FALSE` in the chunk header. It's also much slower than `pairs`.
- On the plus side, it warns you about missing data (if you don't turn that off) and it deals more effectively with factors...

ggpairs Scatterplot Matrix



What Makes a Good Graph? (Tufte, lightly edited)

- During the discovery stage of your work use any style or type of graph you wish. Design becomes important as soon as you want to convey information. At that point you have to create graphs that communicate ideas to others.

What Makes a Good Graph? (Tufte, lightly edited)

- During the discovery stage of your work use any style or type of graph you wish. Design becomes important as soon as you want to convey information. At that point you have to create graphs that communicate ideas to others.
- Graphs communicate most easily when they have a specific message – for instance, "coffee production up!" They lose impact and are less successful when their point is vague – for example, "The number of students in public high schools, 1993-2003."

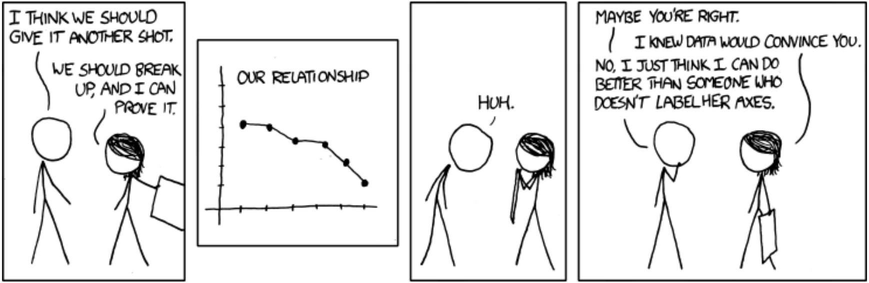
What Makes a Good Graph? (Tufte, lightly edited)

- During the discovery stage of your work use any style or type of graph you wish. Design becomes important as soon as you want to convey information. At that point you have to create graphs that communicate ideas to others.
- Graphs communicate most easily when they have a specific message – for instance, "coffee production up!" They lose impact and are less successful when their point is vague – for example, "The number of students in public high schools, 1993-2003."
- Graphs are powerful when you use the title to reinforce your specific message – "The number of students in public high schools has fallen by a third in ten years." Such transparent messages will be understood and remembered by readers. If you don't tell readers what the graph is saying, some will never know.

What Makes a Good Graph? (Tufte, lightly edited)

- During the discovery stage of your work use any style or type of graph you wish. Design becomes important as soon as you want to convey information. At that point you have to create graphs that communicate ideas to others.
- Graphs communicate most easily when they have a specific message – for instance, "coffee production up!" They lose impact and are less successful when their point is vague – for example, "The number of students in public high schools, 1993-2003."
- Graphs are powerful when you use the title to reinforce your specific message – "The number of students in public high schools has fallen by a third in ten years." Such transparent messages will be understood and remembered by readers. If you don't tell readers what the graph is saying, some will never know.
- After years of hard thinking, I concluded that **graphs are like jokes**: if you have to explain them they have failed.

This doesn't apply to axis labels



Visualizing Categorical Data Well

Building Tables Well

Getting information from a table is like extracting sunlight from a cucumber.

Farquhar AB and Farquhar H

Building Tables Well

There are three key tips related to the development of tables, in practice, as described by Ehrenberg, and also by Howard Wainer[1] who concisely states them as:

1. Order the rows and columns in a way that makes sense.
2. Round - a lot!
3. ALL is different and important.

[1] Visual Revelations (1997), Chapter 10.

Now HERE's a Contingency Table

TABLE 1

Deaths Due to Unexpected Events, by Type of Event, Selected Countries: Mid-1970's

(Rate per 100,000 population)

Country	Year[1]	Deaths due to all causes	Deaths due to unexpected events					
			Total	Transport accidents	Natural factors[2]	Accidents occurring mainly in industry[3]	Homicides and injuries caused intentionally[4]	Other causes[5]
Austria	1975	1,277.2	75.2	34.8	29.7	4.3	1.6	4.8
Belgium	1975	1,218.5	62.6	25.0	25.8	1.5	9	9.4
Canada	1974	742.0	62.1	30.9	18.0	3.9	2.5	6.8
Denmark	1976	1,059.5	41.1	18.3	15.6	1.0	7	5.5
Finland	1974	952.5	62.3	23.7	26.0	2.9	2.6	7.1
France	1974	1,049.5	77.8	23.8	31.0	1.0	9	21.1
Germany (Fed. Rep.)	1975	1,211.8	66.4	24.8	31.6	1.8	1.2	7.0
Ireland	1975	1,060.7	48.6	19.8	20.1	1.9	1.0	5.8
Italy	1974	957.8	47.2	22.8	19.2	1.9	1.1	2.2
Japan	1976	625.6	30.5	13.2	9.7	2.1	1.3	4.2
Netherlands	1975	832.2	40.3	17.8	18.2	1.0	7	2.6
Norway	1976	998.9	48.4	17.3	25.1	1.9	7	3.4
Sweden	1975	1,076.6	55.8	17.2	27.9	1.3	1.1	8.3
Switzerland	1976	904.1	48.4	20.6	20.4	2.1	9	4.4
United Kingdom	1976	1,217.9	34.8	13.0	13.9	1.3	1.1	5.5
United States	1975	888.5	60.6	23.4	15.8	2.6	10.0	8.8

[1] Most current year data available.
[2] Includes fatal accidents due to poisoning, falls, fire, and drowning.
[3] For some countries data relate to accidents caused by machines only.
[4] By another person, including police.
[5] Includes accidents caused by firearms, war injuries, injuries of undetermined causes, and all other accidental causes.

Source: United Nations, World Health Organization, World Health Statistics Annual, 1978, vol. I, Vital Statistics and Cause of Death.

Four Questions

1. What is the general level (per 100,000 population) of accidental death in the countries chosen?
2. How do the countries differ with respect to their rates of accidental death?
3. What are the principal causes of accidental death? Which are the most frequent? The least frequent?
4. Are there any unusual interactions between country and cause of accidental death?

See the Supplementary Table on the Class 11 README page.

Wainer H (1997) *Visual Revelations*, Chapter 10

Deaths due to Unexpected Events, by Type of Event, Selected Countries: Mid-1970's (Rate per 100,000 population)

Country	Year[1]	Deaths due to all causes	Deaths due to unexpected events: Total	Transport Accidents	Natural Factors[2]	Accidents Occurring Mainly in Industry[3]	Homicides and Injuries caused intentionally[4]	Other Causes[5]
Austria	1975	1,277.2	75.2	34.8	29.7	4.3	1.6	4.8
Belgium	1975	1,218.5	62.6	25.0	25.8	1.5	9	9.4
Canada	1974	742.0	62.1	30.9	18.0	3.9	2.5	6.8
Denmark	1976	1,059.5	41.1	18.3	15.6	1.0	7	5.5
Finland	1974	952.5	62.3	23.7	26.0	2.9	2.6	7.1
France	1974	1,049.5	77.8	23.8	31.0	1.0	9	21.1
Germany (Fed. Rep.)	1975	1,211.8	66.4	24.8	31.6	1.8	1.2	7.0
Ireland	1975	1,060.7	48.6	19.8	20.1	1.9	1.0	5.8
Italy	1974	957.8	47.2	22.8	19.2	1.9	1.1	2.2
Japan	1976	625.6	30.5	13.2	9.7	2.1	1.3	4.2
Netherlands	1975	832.2	40.3	17.8	18.2	1.0	7	2.6
Norway	1976	998.9	48.4	17.3	25.1	1.9	7	3.4
Sweden	1975	1,076.6	55.8	17.2	27.9	1.3	1.1	8.3
Switzerland	1976	904.1	48.4	20.6	20.4	2.1	9	4.4
United Kingdom	1976	1,217.9	34.8	13.0	13.9	1.3	1.1	5.5
United States	1975	888.5	60.6	23.4	15.8	2.6	10.0	8.8

Source: United Nations, World Health Organization, World Health Statistics Annual, 1978, vol. I, Vital Statistics and Cause of Death.

Wainer's Three Rules for Table Construction

1. Order the rows and columns in a way that makes sense.
2. Round, a lot!
3. ALL is different and important

- Wainer H (1997) *Visual Revelations* Chapter 10.

Alabama First!

Which is more useful to you?

2013 Percent of Students in grades 9-12 who are obese

State	% Obese	95% CI	Sample Size
Alabama	17.1	(14.6 - 19.9)	1,499
Alaska	12.4	(10.5-14.6)	1,167
Arizona	10.7	(8.3-13.6)	1,520
Arkansas	17.8	(15.7-20.1)	1,470
Connecticut	12.3	(10.2-14.7)	2,270
Delaware	14.2	(12.9-15.6)	2,475
Florida	11.6	(10.5-12.8)	5,491
...			
Wisconsin	11.6	(9.7-13.9)	2,771
Wyoming	10.7	(9.4-12.2)	2,910

or ...

Alabama First!

State	% Obese	95% CI	Sample Size
Kentucky	18.0	(15.7 - 20.6)	1,537
Arkansas	17.8	(15.7 - 20.1)	1,470
Alabama	17.1	(14.6 - 19.9)	1,499
Tennessee	16.9	(15.1 - 18.8)	1,831
Texas	15.7	(13.9 - 17.6)	3,039
...			
Massachusetts	10.2	(8.5 - 12.1)	2,547
Idaho	9.6	(8.2 - 11.1)	1,841
Montana	9.4	(8.4 - 10.5)	4,679
New Jersey	8.7	(6.8 - 11.2)	1,644
Utah	6.4	(4.8 - 8.5)	2,136

It is a rare event when Alabama first is the best choice.

Archiving Data: Sortable Online Tables

2013: Percent of students in grades 9-12 who are obese†

Location Type	Location	Value	95% CI	Sample Size
National	National	13.7	(12.6-14.9)	12580
	Kentucky	18.0	(15.7-20.6)	1537
	Arkansas	17.8	(15.7-20.1)	1470
	Alabama	17.1	(14.6-19.9)	1499
	Tennessee	16.9	(15.1-18.8)	1831
	Texas	15.7	(13.9-17.6)	3039
	West Virginia	15.6	(13.5-18.0)	1561
	Mississippi	15.4	(13.1-17.9)	1446
	Missouri	14.9	(12.3-17.8)	1539
	Delaware	14.2	(12.9-15.6)	2475
	South Carolina	13.9	(11.6-16.5)	1555
	Louisiana	13.5	(11.0-16.4)	1034
	North Dakota	13.5	(11.8-15.3)	1931
	Hawaii	13.4	(11.6-15.4)	4405
	Vermont	13.2	(11.3-15.4)	5853
	Michigan	13.0	(11.4-14.9)	4110
	Ohio	13.0	(10.8-15.5)	1404

Notes on the Data in the previous slides

Source: Estimates from the National Youth Risk Behavior Surveillance System (YRBSS). Available at http://www.cdc.gov/nccdphp/DNPAO/index.html.

To go directly to this table visit this link

- Obese is defined as body mass index (BMI)-for-age and sex $\geq$ 95th percentile based on the 2000 CDC growth chart; BMI was calculated from self-reported weight and height (weight [kg]/ height [m^2]).

Order rows and columns sensibly

- Alabama First!
- Size places - put the largest first. We often look most carefully at the top.
- Order time from the past to the future to help the viewer.
- If there is a clear predictor-outcome relationship, put the predictors in the rows and the outcomes in the columns.

Order the rows and columns sensibly.

Country	Total unexpected deaths	Transport accidents	Natural factors	Industrial accidents	Homicides	Other Causes
France	77.8	23.8	31.0	1.0	0.9	21.1
Austria	75.2	34.8	29.7	4.3	1.6	4.8
Germany	66.4	24.8	31.6	1.8	1.2	7.0
Belgium	62.6	25.0	25.8	1.5	0.9	9.4
Finland	62.3	23.7	26.0	2.9	2.6	7.1
Canada	62.1	30.9	18.0	3.9	2.5	6.8
United States	60.6	23.4	15.8	2.6	10.0	8.8
Sweden	55.8	17.2	27.9	1.3	1.1	8.3
Ireland	48.6	19.8	20.1	1.9	1.0	5.8
Norway	48.4	17.3	25.1	1.9	0.7	3.4
Switzerland	48.4	20.6	20.4	2.1	0.9	4.4
Italy	47.2	22.8	19.2	1.9	1.1	2.2
Denmark	41.1	18.3	15.6	1.0	0.7	5.5
Netherlands	40.3	17.8	18.2	1.0	0.7	2.6
United Kingdom	34.8	13.0	13.9	1.3	1.1	5.5
Japan	30.5	13.2	9.7	2.1	1.3	4.2

Round - a lot!

- Humans cannot understand more than two digits very easily.
- We almost never care about accuracy of more than two digits.
- We can almost never justify more than two digits of accuracy statistically.

Suppose we want to report a correlation coefficient of 0.25

- How many observations do you think you would need to justify such a choice?
- To report 0.25 meaningfully, we should know the second digit isn't 4 or 6, right?

Reporting a correlation coefficient of 0.25

To report 0.25 meaningfully, we desire to be sure that the second digit isn't 4 or 6.

- That requires a standard error less than 0.005
- The *standard error* of any statistic is proportional to 1 over the square root of the sample size, n.

So $\frac{1}{\sqrt{n}} \sim 0.005$, but that means $\sqrt{n} = \frac{1}{0.005} = 200$.

And if $\sqrt{n} = 200$, then $n = (200)^2 = 40{,}000$.

Do we usually have 40,000 observations?

Round, a lot!

Country	Total unexpected deaths	Transport accidents	Natural factors	Industrial accidents	Homicides	Other Causes
France	78	24	31	1	1	21
Austria	75	35	30	4	2	5
Germany	66	25	32	2	1	7
Belgium	63	25	26	2	1	9
Finland	62	24	26	3	3	7
Canada	62	31	18	4	3	7
United States	61	23	16	3	10	9
Sweden	56	17	28	1	1	8
Ireland	49	20	20	2	1	6
Norway	48	17	25	2	1	3
Switzerland	48	21	20	2	1	4
Italy	47	23	19	2	1	2
Denmark	41	18	16	1	1	6
Netherlands	40	18	18	1	1	3
United Kingdom	35	13	14	1	1	6
Japan	31	13	10	2	1	4

ALL is different and important

Country	Total unexpected deaths	Transport accidents	Natural factors	Industrial accidents	Homicides	Other Causes
France	78	24	31	1	1	21
Austria	75	35	30	4	2	5
Germany	66	25	32	2	1	7
Belgium	63	25	26	2	1	9
Finland	62	24	26	3	3	7
Canada	62	31	18	4	3	7
United States	61	23	16	3	10	9
Sweden	56	17	28	1	1	8
Ireland	49	20	20	2	1	6
Norway	48	17	25	2	1	3
Switzerland	48	21	20	2	1	4
Italy	47	23	19	2	1	2
Denmark	41	18	16	1	1	6
Netherlands	40	18	18	1	1	3
United Kingdom	35	13	14	1	1	6
Japan	31	13	10	2	1	4

Cluster when you can, and highlight outliers.

Country	Total unexpected deaths	Transport accidents	Natural factors	Industrial accidents	Homicides	Other Causes
France	**78**	24	31	1	1	**21**
Austria	**75**	**35**	30	4	2	5
Germany	**66**	25	32	2	1	7
Belgium	**63**	25	26	2	1	9
Finland	**62**	24	26	3	3	7
Canada	**62**	**31**	**18**	4	3	7
United States	**61**	23	**16**	3	**10**	9
Sweden	**56**	17	28	1	1	8
Ireland	**49**	20	20	2	1	6
Norway	**48**	17	25	2	1	3
Switzerland	**48**	21	20	2	1	4
Italy	**47**	23	19	2	1	2
Denmark	**41**	18	16	1	1	6
Netherlands	**40**	18	18	1	1	3
United Kingdom	**35**	13	14	1	1	6
Japan	**31**	13	10	2	1	4

Visualizing Categories

http://flowingdata.com/projects/2016/alcohol-world/

Recorded APC is defined as the recorded amount of alcohol consumed per capita (15+ years) over a calendar year in a country, in litres of pure alcohol. The indicator only takes into account the consumption which is recorded from production, import, export, and sales data often via taxation.

- Numerator: The amount of recorded alcohol consumed per capita (15+ years) during a calendar year, in litres of pure alcohol.
- Denominator: Midyear resident population (15+ years) for the same calendar year, UN World Population Prospects, medium variant.

http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=462

Elements of Data Analytic Style

Leek, Chapter 3 (Tidying the Data)

Components of a Processed Data Set

1. The raw data.
2. A tidy data set.
3. A code book describing each variable and its values in the tidy data set.
4. An explicit and exact recipe you used to go from 1 to 2 to 3.

See https://github.com/jtleek/datasharing for a guide for your project.

Tidy Data Video from Hadley Wickham https://vimeo.com/33727555

Leek, Chapter 4 (Checking the Data)

- Coding variables appropriately
  - Continuous, Ordinal, Categorical, Missing, Censored
- Code categorical / ordinal variables so that R will read them as factors.
- Encode everything using text, not with colors on the spreadsheet.
- Identify the missing value indicator, and use `NA` whenever you can.
- Check for coding errors, particularly label switching.

Leek, Chapter 12 (Reproducibility)

Reproducibility of workflow is what we're aiming for.

- Everything in a script. (R Markdown)
- Everything stored in a plain text file (future-proof: .csv, .Rmd)
- Organize your data analysis in subfolders of the project directory
- Use version control (something I should do more of)
- Add `sessionInfo()` command to final version of work when you need to preserve the details on software and parameters - see next slide.

My session info, at home, 2018-10-01

Include this information in your project submissions, but not probably in your other assignments, unless we ask you for it.

```
> sessionInfo()
R version 3.5.1 (2018-07-02)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows >= 8 x64 (build 9200)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252    LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C                           LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] bindrcpp_0.2.2  forcats_0.3.0   stringr_1.3.1   dplyr_0.7.6     purrr_0.2.5     readr_1.1.1     tidyr_0.8.1     tibble_1.4.2
 [9] tidyverse_1.2.1 GGally_1.4.0    ggplot2_3.0.0

loaded via a namespace (and not attached):
 [1] ggrepel_0.8.0      Rcpp_0.12.18       lubridate_1.7.4    lattice_0.20-35    prettyunits_1.0.2  assertthat_0.2.0   rprojroot_1.3-2
 [8] digest_0.6.15      packrat_0.4.9-3    R6_2.2.2           cellranger_1.1.0   plyr_1.8.4         backports_1.1.2    evaluate_0.11
[15] ggstance_0.3.1     httr_1.3.1         highr_0.7          pillar_1.3.0       rlang_0.2.2        progress_1.2.0     lazyeval_0.2.1
[22] readxl_1.1.0       rstudioapi_0.7     Matrix_1.2-14      rmarkdown_1.10     labeling_0.3       splines_3.5.1      munsell_0.5.0
[29] broom_0.5.0        compiler_3.5.1     modelr_0.1.2       pkgconfig_2.0.2    htmltools_0.3.6    tidyselect_0.2.4   gridExtra_2.3
[36] mosaicCore_0.6.0   reshape_0.8.7      viridisLite_0.3.0  crayon_1.3.4       withr_2.1.2        MASS_7.3-50        grid_3.5.1
[43] nlme_3.1-137       mosaicData_0.17.0  jsonlite_1.5       gtable_0.2.0       ggformula_0.9.0    magrittr_1.5       scales_1.0.0
[50] cli_1.0.0          stringi_1.2.4      reshape2_1.4.3     xml2_1.2.0         ggdendro_0.1-20    RColorBrewer_1.1-2 tools_3.5.1
[57] glue_1.3.0         hms_0.4.2          yaml_2.2.0         colorspace_1.3-2   mosaic_1.4.0       rvest_0.3.2        knitr_1.20
[64] bindr_0.1.1        haven_1.1.2
```

My session info, at home, One Year Ago

Here is the 2017-10-03 version of this information. At the time R 3.4.2 was brand new!

```
> sessionInfo()
R version 3.4.2 (2017-09-28)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows >= 8 x64 (build 9200)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252    LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C                           LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] bindrcpp_0.2    dplyr_0.7.4     purrr_0.2.3     readr_1.1.1     tidyr_0.7.1     tibble_1.3.4    ggplot2_2.2.1   tidyverse_1.1.1
[9] GGally_1.3.2

loaded via a namespace (and not attached):
 [1] progress_1.1.2     reshape2_1.4.2     haven_1.1.0        lattice_0.20-35    colorspace_1.3-2   htmltools_0.3.6    yaml_2.1.14
 [8] rlang_0.1.2        foreign_0.8-69     glue_1.1.1         RColorBrewer_1.1-2 modelr_0.1.1       readxl_1.0.0       bindr_0.1
[15] plyr_1.8.4         stringr_1.2.0      munsell_0.4.3      gtable_0.2.0       cellranger_1.1.0   rvest_0.3.2        psych_1.7.8
[22] evaluate_0.10.1    labeling_0.3       knitr_1.17         forcats_0.2.0      parallel_3.4.2     highr_0.6          broom_0.4.2
[29] Rcpp_0.12.13       scales_0.5.0       backports_1.1.1    jsonlite_1.5       mnormt_1.5-5       hms_0.3            digest_0.6.12
[36] stringi_1.1.5      grid_3.4.2         rprojroot_1.2      tools_3.4.2        magrittr_1.5       lazyeval_0.2.0     pkgconfig_2.0.1
[43] prettyunits_1.0.2  xml2_1.1.1         lubridate_1.6.0    assertthat_0.2.0   rmarkdown_1.6      reshape_0.8.7      httr_1.3.1
[50] R6_2.2.2           nlme_3.1-131       compiler_3.4.2
```