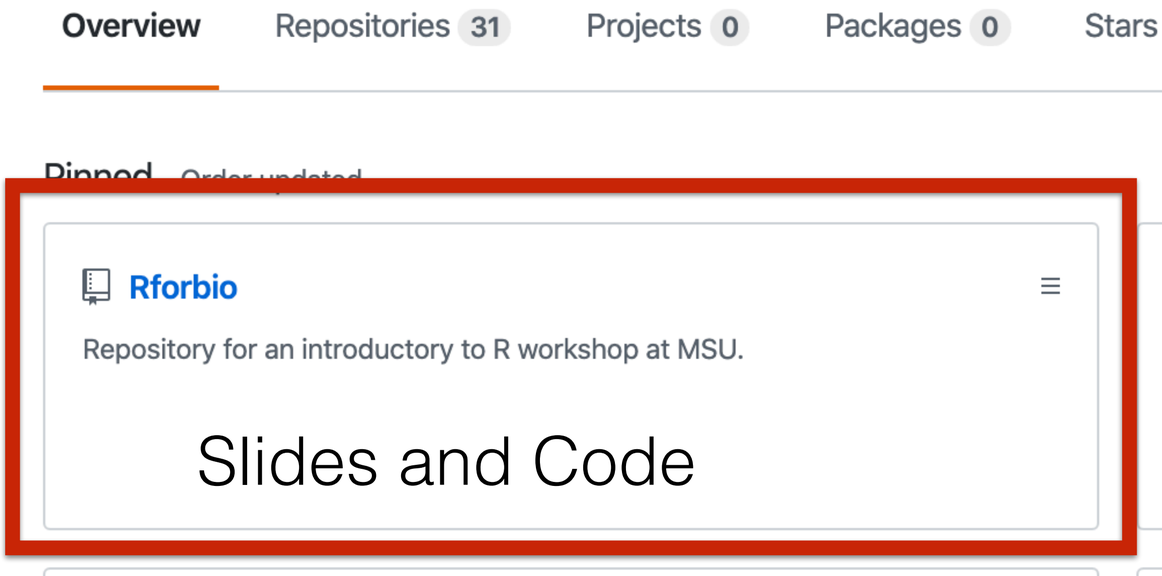
Please Download

cran.r-project.org/

rstudio.com

github.com/nguyens7/



Introduction to R Workshop

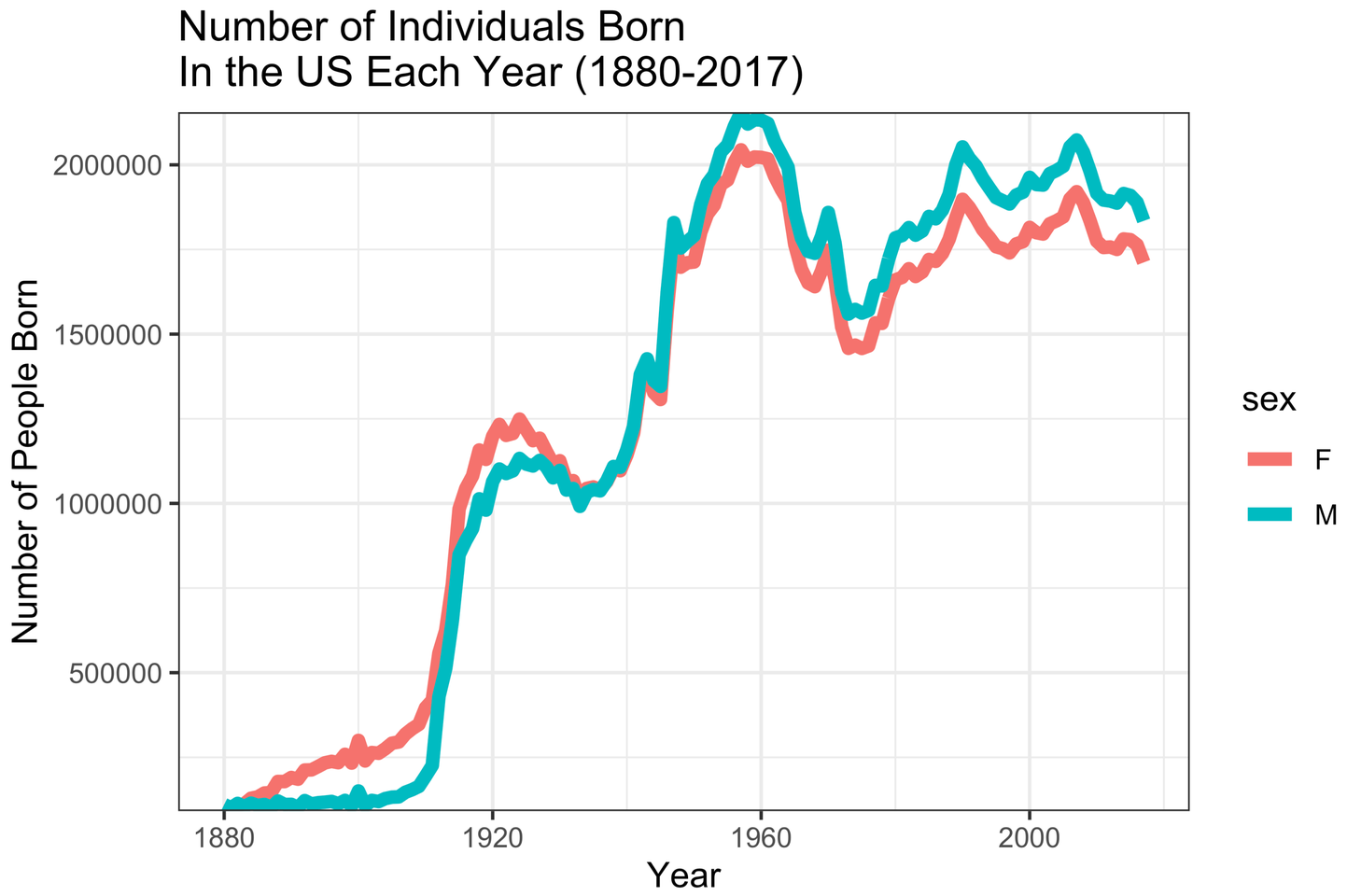
CMB/GGS Seminar

Sean Nguyen

February 12, 2020

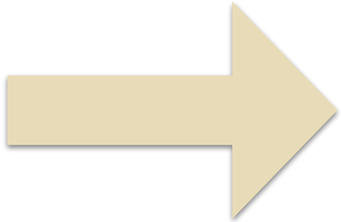
Today's Goals

- **Install** R and Rstudio
- **Import** packages
- **Explore** a dataset
- **Visualize** data



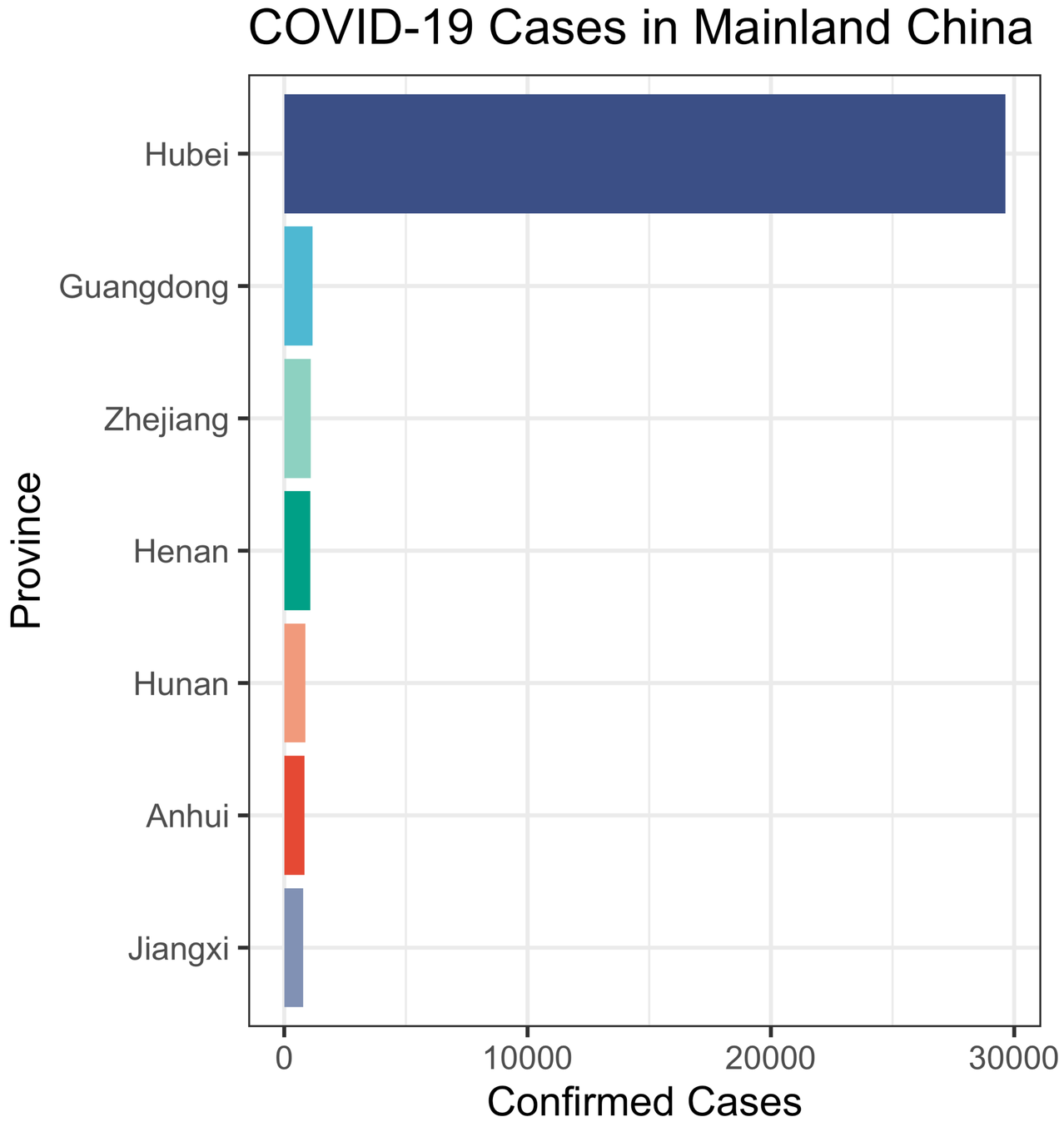
Programming language for statistical computing

R is good for: reproducible analysis

Raw data  **Analyzed** data

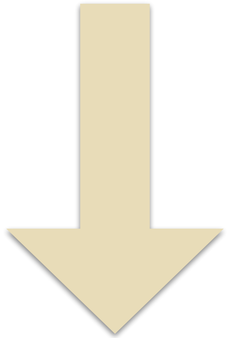
R is good for:
generating
beautiful figures

Raw data



R is good for: calculating statistics

Raw data



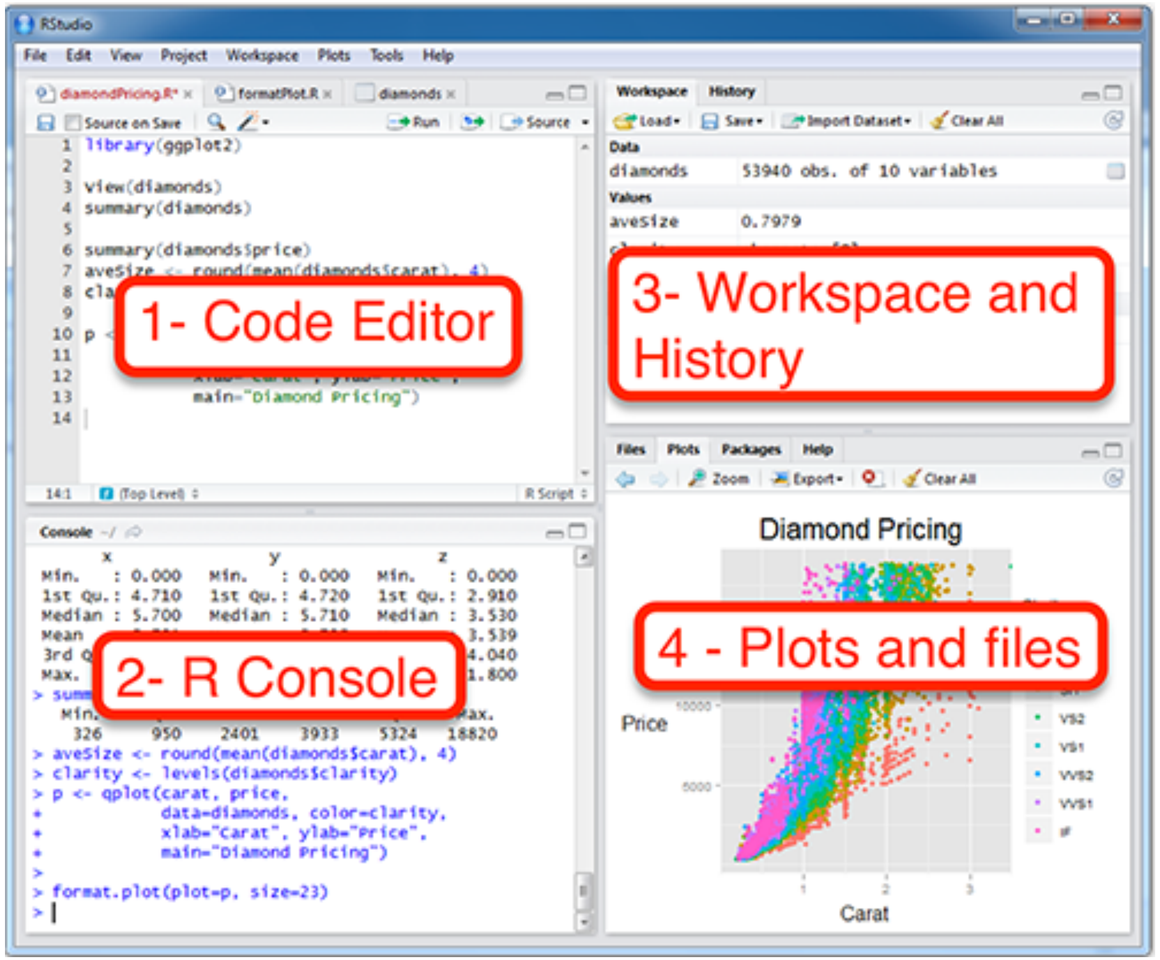
```
ANOVA <- aov(mean~(Organism*Treatment),data=data4)
tidy(ANOVA)
```

term <chr>	df <dbl>	sumsq <dbl>	meansq <dbl>	statistic <dbl>	p.value <dbl>
Organism	2	26807.853	13403.9267	43.48849	3.176012e-06
Treatment	1	21687.502	21687.5022	70.36422	2.306759e-06
Organism:Treatment	2	16466.031	8233.0156	26.71168	3.807941e-05
Residuals	12	3698.613	308.2178	*NA*	*NA*

4 rows

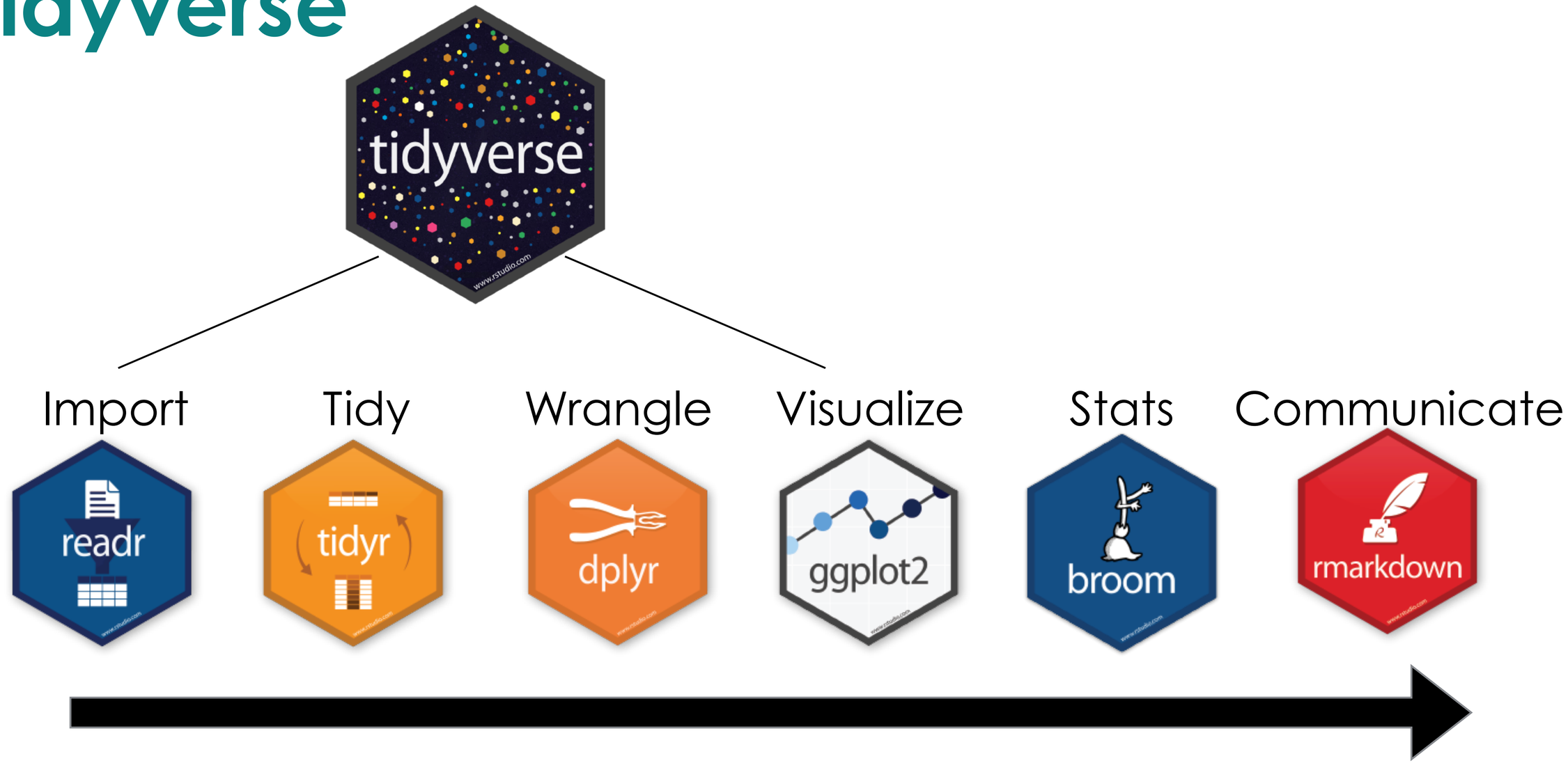
RStudio®

Integrated development environment (IDE) for easy creation and organization of R scripts



Packages are: a collection of useful functions

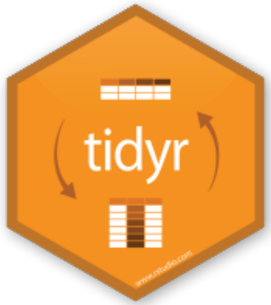
Data Analysis in the **Tidyverse**



Data Analysis in the **Tidyverse**

Import

Tidy



Wrangle

filter()
group_by()
arrange()
summarise()

Visualize

ggplot()
geom_line()
geom_bar()
geom_point()

Stats

Communicate

Install packages

Package - Collection of R functions

- Only install once
- Load them each time you run a script

tidyverse, babynames

	year	sex	name	n	prop
1	1880	F	Mary	7065	0.0723843285
2	1880	F	Anna	2604	0.0266792345
3	1880	F	Emma	2003	0.0205216999
4	1880	F	Elizabeth	1939	0.0198659891
5	1880	F	Minnie	1746	0.0178886111
6	1880	F	Margaret	1578	0.0161673702
7	1880	F	Ida	1472	0.0150813491
8	1880	F	Alice	1414	0.0144871112
9	1880	F	Bertha	1320	0.0135240359
10	1880	F	Sarah	1288	0.0131961805
11	1880	F	Annie	1258	0.0128888160
12	1880	F	Clara	1226	0.0125609606

The Assignment Operator

- Assigns value to an **object**

<-

```
x <- 4
x
> 4
```

Pipe operator %>%

- **Interpreted as "then"**

Fruit	Count
Apple	34
Raspberry	67
Pear	35
Plum	27
Peach	5
Strawberry	2
Melon	97
Mango	5

```
data %>%
    filter(Fruit == "Raspberry")
```

Fruit	Count
Raspberry	67

dplyr - clean up/aggregate data

- filter()
- arrange()
- group_by()
- summarize()

filter()- picks **rows** based on values

Fruit	Count
Apple	34
Raspberry	67
Pear	35
Plum	27
Peach	5
Strawberry	2
Melon	97
Mango	5

filter(Fruit == “Raspberry”)

Fruit	Count
Raspberry	67

filter(Count < 10)

Fruit	Count
Peach	5
Strawberry	2
Mango	5

arrange()- changes row order

Fruit	Count
Apple	34
Raspberry	67
Pear	35
Plum	27
Peach	5
Strawberry	2
Melon	97
Mango	5

arrange(desc(Count)

Fruit	Count
Melon	97
Raspberry	67
Pear	35
Apple	34
Mango	5
Peach	5

- **group_by()**- '**lock-in**' by certain criteria
- **summarize()** - **reduce** multiple values to a **single value**

Cat	Fruit	Count
1	Apple	34
1	Raspberry	67
1	Pear	35
1	Plum	27
2	Peach	5
2	Strawberry	2
2	Melon	97
2	Mango	5

```
data %>%
    group_by(Cat) %>%
    summarize( Total = sum(Count))
```

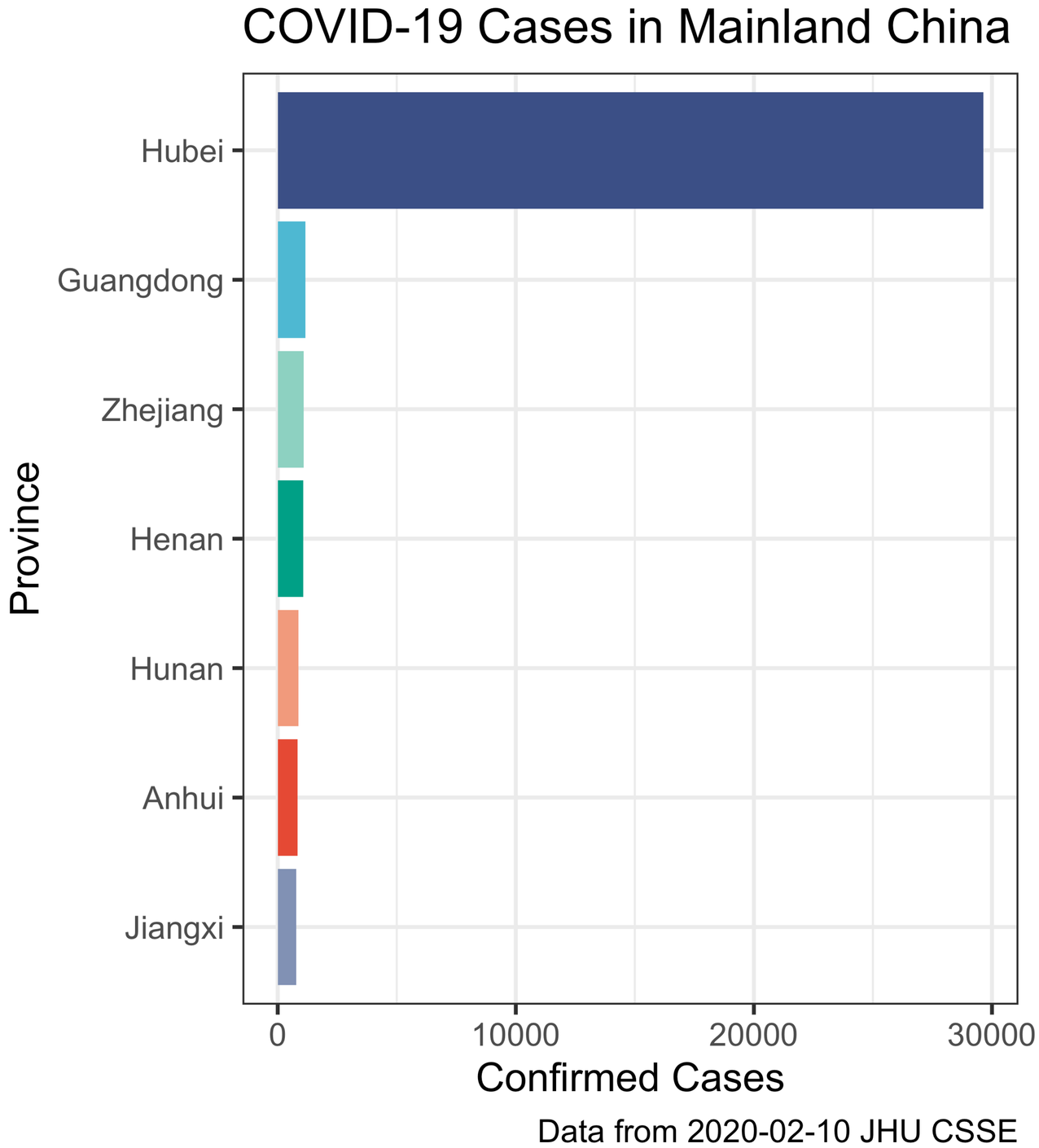
Cat	Total
1	163
2	109

dplyr - clean up/aggregate data

- **filter()**- picks **rows** based on values
- **arrange()**- changes **row order**
- **group_by()**- '**lock-in**' by certain criteria
- **summarize()** - **reduce** multiple values to a **single value**

ggplot2

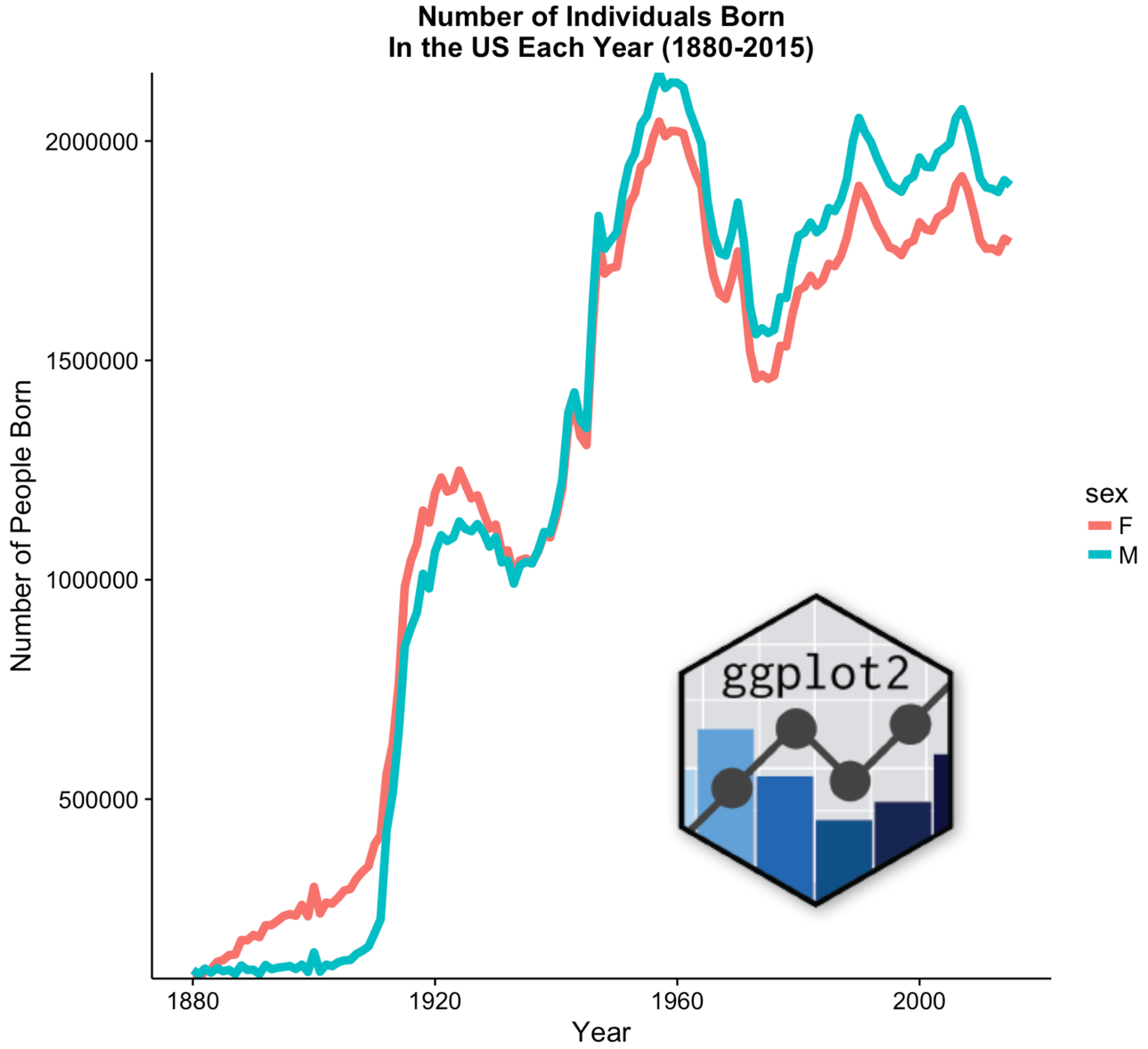
Powerful graphing package for generating high quality figures based on the grammar of graphics



ggplot2

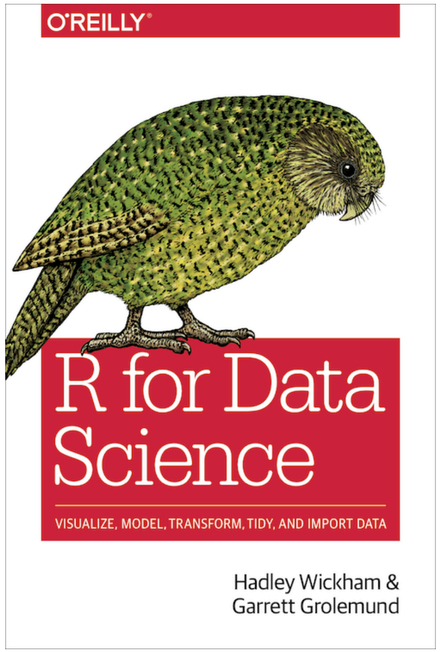
```
data %>%
    ggplot(aes( x =  year,
                y = n,
                color = sex) +
      geom_line()
```

	year	sex	name	n	prop
1	1880	F	Mary	7065	0.0723843285
2	1880	F	Anna	2604	0.0266792345
3	1880	F	Emma	2003	0.0205216999
4	1880	F	Elizabeth	1939	0.0198659891
5	1880	F	Minnie	1746	0.0178886111
6	1880	F	Margaret	1578	0.0161673702



Resources

YouTube



GitHub

stackoverflow

Demo!