

Project 1

Prof. Roger D. Peng

*Department of Statistics and Data Sciences
University of Texas at Austin*

Spring 2024

Overview

- **Goal:** Apply the skills you have learned so far to ask and answer data-related questions
 - Tidyverse: tidy data, ggplot2, data wrangling and transformation, exploratory data analysis
- **Final product:** A PDF report containing your code and results (see format in the instructions)
- Can work independently or in a group of at most 3 people (requirements are different)
- Groups do **not** have to be the same as Lab Groups

Speedcubing

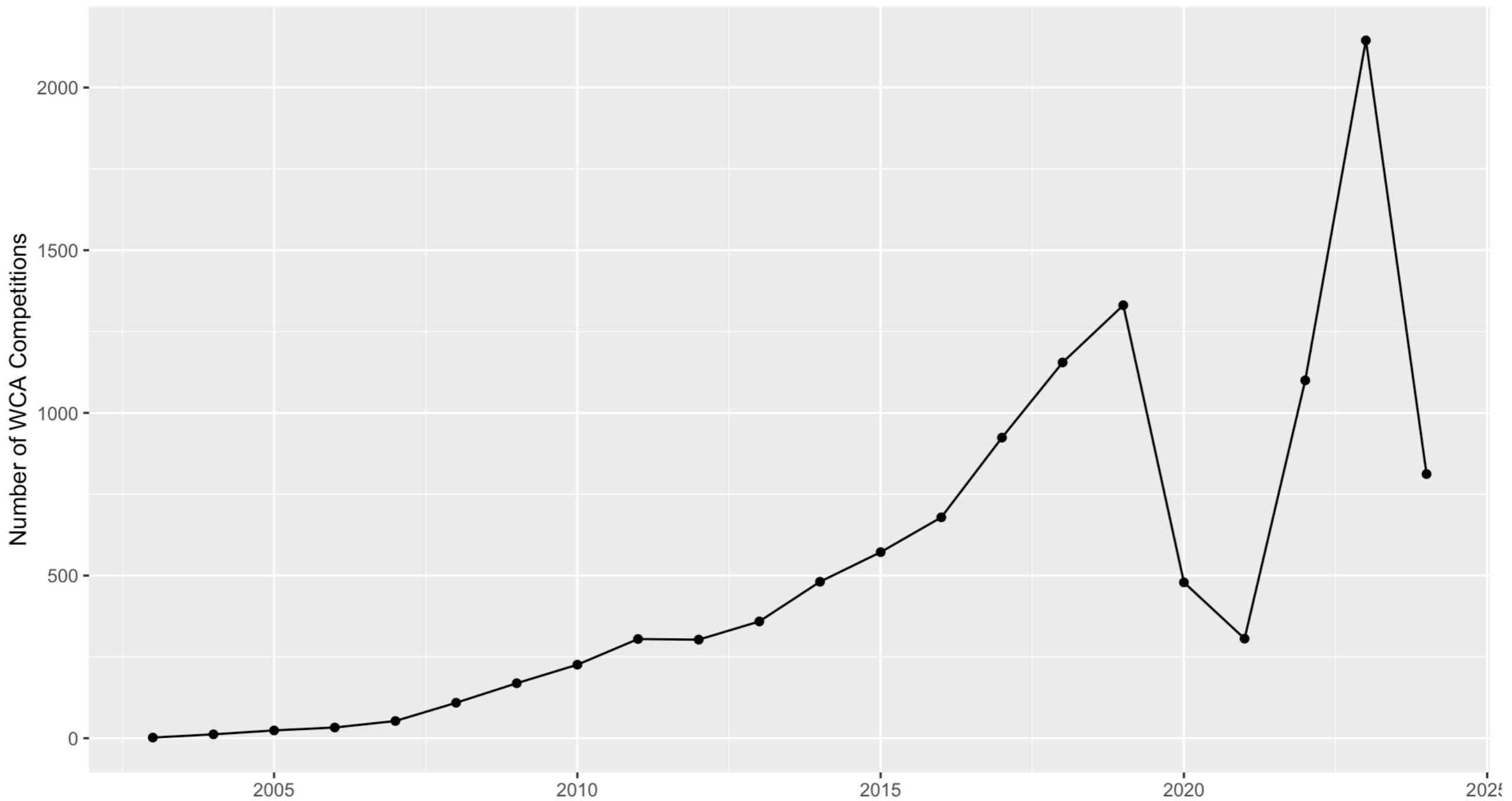
- Original Rubik's Cube invented in 1974 by Erno Rubik, a Hungarian inventor.
- First official speedcubing competition held in 1982 in Budapest
 - Winning single solve time was 22.95 seconds
- Second official speedcubing competition held in 2003 in Toronto
- Many speedcube manufacturers (not Rubik's original)

Speedcubing



Speedcubing Competitions

WCA Speedcubing Competitions By Year, 2003--2024





<https://youtu.be/kOfh3iZD00k>

Dana Yi



https://youtu.be/gh8HX4itF_w

Max Park

Speedcubing

- Main event is usually the 3x3x3 cube
- There are many other events in a given competition
- Format for each round is "average of 5" which uses a trimmed mean (dropping the highest and lowest times)
- Everyone in a round gets the same scramble so differences in time in a round are due to solving skill

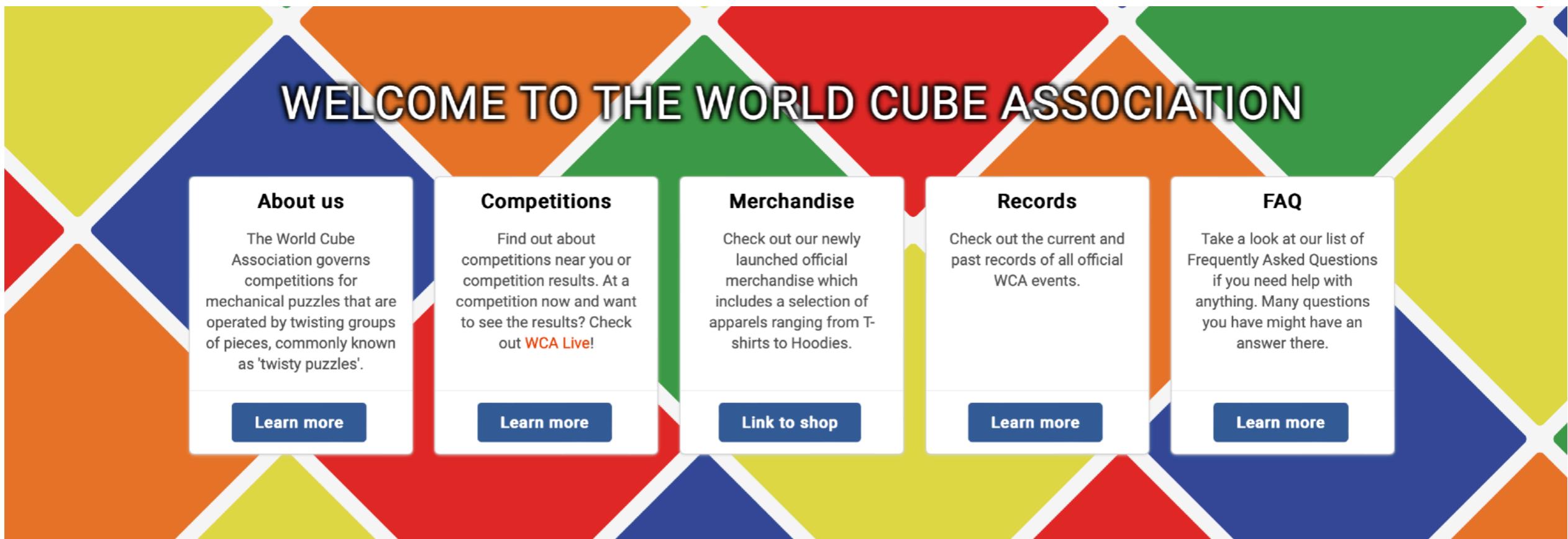
Cube Solving: CFOP

- Developed in the 1980s and formalized by Jessica Fridrich in 1997
 - Cross
 - First two layers (F2L)
 - Orientation of the last layer (OLL)
 - Permutation of the last layer (PLL)
- Many variations

Why Speedcubing?

- Complete data with > 20 year history available for free and updated in real-time
- Relatively easy to understand and learn about
- Non-competitive; not team-based
- Not particularly popular compared to other sports/games

World Cube Association



Latest news

Continental Championships 2024 Announcement - Oceanic, Asia & Africa (February 2024)

Announced by Ram Thakkar on Monday, February 12, 2024 at 2:26 AM CST

We are happy to announce that **Perth, Western Australia** will be hosting the 2024 WCA Oceanic Championship, **Kuala Lumpur, Malaysia** will be hosting the 2024 WCA Asian Championship and **Cape Town, South Africa** will be hosting the 2024 WCA African Championship.

We would also like to thank Astana, Kazakhstan for applying for the WCA Asian Championship 2024.

You can view the application documents [here](#) (WCA Asian Championship), [here](#) (WCA Oceanic Championship) and [here](#) (WCA African Championship).

Announcements

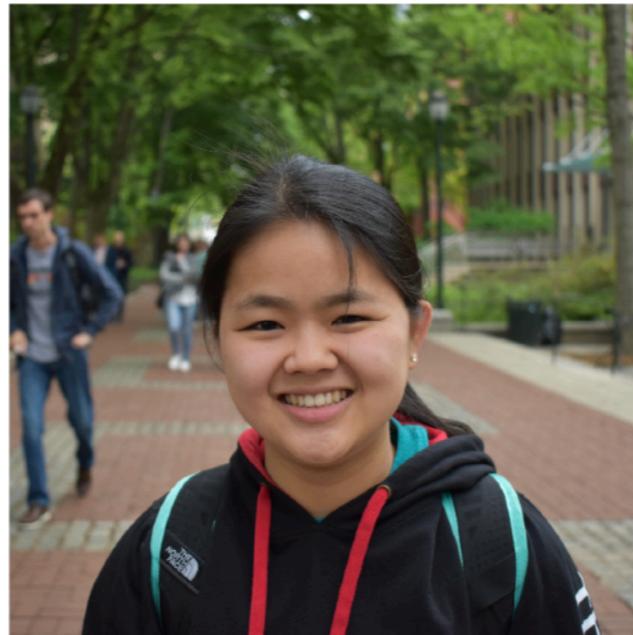
- [Continental Championships 2024 Announcement - Oceanic, Asia & Africa \(February 2024\)](#)
- [WDC Leader Announcement \(February 2024\)](#)
- [Changes to the WCA Board of Directors \(January 2024\)](#)
- [Call for Applicants - WMT Members \(January 2024\)](#)
- [Call for Applicants - WRC Members \(January 2024\)](#)
- [WRT Leader Announcement \(January 2024\)](#)
- [Call for Applicants - WCT Members \(January 2024\)](#)
- [World Cube Association Hosts "Newcomer Month," Inviting Everyone to Join the Thrilling World of Competitive Speedcubing](#)
- [Year End Message 2023](#)
- [WDC Digest 2023-Q4](#)

Data

- The dataset comes from the World Cube Association database and contains times / scores of individual speed cubers from all over the world
- Data are divided into 13 separate tables containing different information about people, competitions, and results.
- Data have been limited to 3x3x3 events to make the dataset a more reasonable size
- The data are relatively clean but that does not mean that you won't have to do some transformation

Dana Yi

Delegate



Region	WCA ID	Gender	Competitions	Completed Solves
United States	2010YIDA01	Female	109	4889

Current Personal Records

Event	NR	CR	WR	Single	Average	WR	CR	NR
3x3x3 Cube	42	49	133	4.98	7.23	358	118	97
2x2x2 Cube	1151	1455	5256	2.00	3.10	2973	834	654
4x4x4 Cube	103	120	411	26.37	30.64	375	103	92
5x5x5 Cube	58	68	240	49.34	55.79	256	70	61
6x6x6 Cube	49	59	180	1:32.81	1:39.77	166	51	42
7x7x7 Cube	47	57	212	2:26.44	2:32.68	172	42	37
3x3x3 One-Handed	308	384	1343	12.21	17.12	2383	612	474
Clock	300	340	1106	6.50	8.21	1202	354	310
Megaminx	76	94	385	42.39	47.56	347	86	68
Pyraminx	634	790	3087	3.11	5.90	4853	1190	932
Skewb	1662	2131	7753	4.74	7.59	7669	2054	1566
Square-1	425	510	1406	11.44	19.39	2379	822	667

Max Park



Region	WCA ID	Gender	Competitions	Completed Solves				
United States	2012PARK03	Male	166	5954				
Current Personal Records								
Event	NR	CR	WR	Single	Average	WR	CR	NR
3x3x3 Cube	1	1	1	3.13	4.86	2	1	1
2x2x2 Cube	3374	4461	15887	2.88	4.31	12020	3420	2614
4x4x4 Cube	1	1	1	16.79	19.38	1	1	1
5x5x5 Cube	1	1	1	32.60	35.94	1	1	1
6x6x6 Cube	1	1	1	59.74	1:07.11	2	1	1
7x7x7 Cube	1	1	1	1:35.68	1:41.78	1	1	1
3x3x3 One-Handed	1	1	1	6.20	8.62	1	1	1
Square-1	3742	4698	14248	52.69	1:10.20	12639	4172	3302

Data

Rankings

Last updated: Wednesday, February 21, 2024 at 9:14 AM CST

Please note that some filters (such as past years) are temporarily disabled. The Software Team is working to restore these filters - we are sorry for the inconvenience.

Event	Region											
Type	Years	Gender	Show	World								
Single	Average	All years	Only	Until	All	Male	Female	100 persons	Results	By region		

#	Name	Result	Representing	Competition
1	Max Park	3.13	🇺🇸 United States	Pride in Long Beach 2023
2	Luke Garrett	3.44	🇺🇸 United States	Flag City Summer 2023
3	Yiheng Wang (王艺衡)	3.47	🇨🇳 China	MYHM Singapore Championship 2024
3	Yusheng Du (杜宇生)	3.47	🇨🇳 China	Wuhu Open 2018
5	Tymon Kolasiński	3.67	🇵🇱 Poland	Melbourne Summer 2024
6	Jode Brewster	3.88	🇦🇺 Australia	Tassie Summer 2023
7	Asher Kim-Magierek	3.89	🇺🇸 United States	Rose City 2022
8	Ruihang Xu (许瑞航)	4.01	🇨🇳 China	Vietnam Championship 2023
9	Natthaphat Mahtani (ณัฐภัทร จิ มหาทานี)	4.02	🇹🇭 Thailand	Bangkok Cube Day Winter 2024
10	Max Siauw	4.03	🇺🇸 United States	BC Cubing Springback A 2022
11	Matty Hiroto Inaba	4.05	🇺🇸 United States	Honolulu Li'ili'i 2024
11	Radosław Marcinek	4.05	🇵🇱 Poland	SLS Październik 2023
13	Ianis Costin Chele	4.07	🇷🇴 Romania	Bucharest Cubing February B 2024
14	Sean Patrick Villanueva	4.11	🇵🇭 Philippines	Valenzuela Cubing Open 2023
15	Brennen Lin	4.14	🇨🇦 Canada	SOS Oshawa 2023

Data

Rankings

Last updated: Wednesday, February 21, 2024 at 9:14 AM CST

Please note that some filters (such as past years) are temporarily disabled. The Software Team is working to restore these filters - we are sorry for the inconvenience.

Event		Region									
Type	Years	Gender			Show						
Single	Average	All years	Only	Until	All	Male	Female	100 persons	Results	By region	
World											
#	Name	Result	Representing	Competition					Solves		
1	Yiheng Wang (王艺衡)	4.48	🇨🇳 China	🇸🇬 Mofunland Cruise Open 2023		4.72	4.72	3.99	(3.95)	(5.99)	
2	Max Park	4.86	🇺🇸 United States	🇺🇸 Marshall Cubing September 2022		4.62	4.78	(5.68)	5.19	(4.50)	
2	Tymon Kolasiński	4.86	🇵🇱 Poland	🇵🇱 Cube4fun Warsaw 2022		(4.02)	4.68	5.33	4.56	(5.59)	
4	Ruihang Xu (许瑞航)	5.01	🇨🇳 China	🇻🇳 Vietnam Championship 2023		(4.01)	4.87	4.60	(6.12)	5.56	
5	Luke Garrett	5.03	🇺🇸 United States	🇺🇸 Pittsburgh Fall 2022		4.66	5.40	(4.47)	(6.25)	5.02	
6	Matty Hiroto Inaba	5.10	🇺🇸 United States	🇺🇸 Hawai'i Big Island Winter 2024		5.13	(5.70)	4.95	(4.60)	5.21	
7	Yufang Du (杜昱方)	5.15	🇨🇳 China	🇨🇳 Hefei Winter Open 2023		5.50	4.41	5.55	(7.72)	(4.31)	
8	Leo Borromeo	5.24	🇵🇭 Philippines	🇵🇭 Cebu New Year Open 2023		5.29	5.47	(6.72)	4.97	(4.31)	
9	Twan Dullemond	5.44	🇳🇱 Netherlands	🇩🇪 Bavarian Open 2023		(4.81)	5.00	5.80	(7.44)	5.51	
10	Luke Griesser	5.51	🇺🇸 United States	🇺🇸 Mason Cube Challenge 2023		5.40	(7.20)	5.83	5.29	(5.12)	
11	Feliks Zemdegs	5.53	🇦🇺 Australia	🇦🇺 Odd Day in Sydney 2019		(7.16)	5.04	(4.67)	6.55	4.99	
12	Timofei Tarasenko	5.54	🇷🇺 Russia	🇬🇪 Caucasus Open 2023		5.17	(8.37)	6.02	5.43	(5.10)	
13	Bofan Zhang (张博藩)	5.55	🇨🇳 China	🇨🇳 Chengdu Flash Cube Open 2023		5.48	5.76	(6.22)	(5.26)	5.40	
14	Patrick Ponce	5.57	🇺🇸 United States	🇺🇸 East Brunswick Open 2022		5.60	5.82	(9.16)	(4.96)	5.29	
14	Yezhen Han (韩业臻)	5.57	🇨🇳 China	🇨🇳 Guangdong Open 2021		5.87	5.42	(5.30)	(7.53)	5.42	

Required Questions

1. How many active (3x3x3) speed cubers are there registered with the WCA? (active cuber = competed in at least two competitions in 2022--2024)
2. (a) Who holds the current world record single? (b) Who previously held the world record single? When were each of these records set?
3. Who is the top ranked male speed cuber (single) in Australia? Who is the top ranked female (single) speed cuber in Europe?
4. For the top 10 speed cubers in the world, on average, how many solves do they have to do before achieving a sub-5 second solve?
5. Which speed cubers *not* ranked in the top 10,000 (worldwide for single best) should I keep an eye on for the near future?

Additional Questions

- You must develop **two** additional questions on your own that you can explore with the data (working individually)
- If you are working in a **group** you must develop **three** additional questions (as a group)
- Each of the questions **must** require you to join at least two tables (i.e. you have to do at least one join)

Types of Questions

- **Basic Descriptive Questions**
 - How many speedcubers are there in the USA?
 - Who are the top ranked speedcubers right now?
- **Current Trends Questions**
 - Are Europeans getting better or worse relative to the USA?
 - Who are the up-and-comers? Who should we be watching out for?
- **Speculative Questions**
 - Which country will dominate cubing five years from now?
 - Who will be ranked number 1 two years from now?

Report Sections

- Describe 2-3 additional questions that you developed and why you chose them (groups do 3 questions)
- Answer Questions 1-5
- Answer your 2-3 additional questions
- Discussion and Reflection; compare expectations to observed data

Format

- Create the report using R Markdown with headers for each section
- Include comments to the R code
- Include references (datasets, context)
- The final report should be no more than 20 pages including all code/graphics/output (the number of pages can vary greatly depending on the cleaning process)
- Submission on Gradescope

Two Parts

- Part 1
 - Develop your additional questions
 - State expectations for all additional questions
 - Develop a detailed analysis plan for one of your additional questions (review Lecture 4 "Basic Structure of Data Analyses")
- Part 2
 - Analyze the data
 - Write and submit your final report