

CIF21 DIBBs: EI: Vizier, Streamlined Data Curation







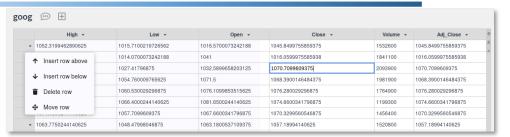
PI: Oliver Kennedy¹, Co-Pis: Juliana Freire², Boris Glavic³

Institutions: ¹ University at Buffalo, ² New York University, ³ Illinois Institute of Technology

 Scalable open-source platform for data debugging and exploration

https://github.com/VizierDB

- Combines notebooks and spreadsheets for ease of use
- Captures provenance both for data and computations
 - Reproducibility, re-use, and transparency
 - Automated versioning
 - Tracking of potential data errors
- Vizier(Dirty Data) = Better Data + Provenance
 - Publish data with provenance and known caveats



		ew_lork_city_country_course	s_of_Death_12_11_2018.csv	
Q	causes (YEAR(ant), LEADING CAUSE(varchar), SEX(varchar), RACE ETHNICLITY(varchar), DEATHS (ant), DEATH RATE(real), AGE ADJUSTED_DEATH_RATE(real)) 1994 row(s)			
		+		
[2]	1 SELECT sex, SUM(deaths) FROM causes GROUP BY sex			
	Output Dataset by_gender			
⊞ '	by_gender			
	-7_5-11-11	SEX	SUM	
	0 M		207927	
	1 F		217071	
	1 F		217071	
			217071	
			217071	
		+	217071	
		<u>+</u>	217071	
[3]		gender	217071	
	₩ < > № ± €	gender	217071	
₫ Ву	CREATE PLOT 'By Gender' FOR by_		217071	
₫ Ву	HM < > > >		217071	
₫ Ву	CREATE PLOT 'By Gender' FOR by_	- H	217071	
₫ Ву	CREATE PLOT 'By Gender' FOR by_	gender	217071	
	CREATE PLOT 'By Gender' FOR by_ Gender Bar Chart			
Ы Ву	CREATE PLOT 'By Gender' FOR by_ Gender harts Bar Chart Deaths		Deaths	
Ы Ву	CREATE PLOT 'By Gender' FOR by_ Gender Bar Chart			