

CU Jail Data

This data set is a subset of the Champaign County Sheriff Office data on bookings throughout the county for the year 2012. It contains no personal identifying information for people who were booked, but does include demographic variables and details about the crime and category of crime. The data set is a comma separated values (.csv) file called `ccso_subset_2` and contains 2078 rows and 11 columns.

CU Jail Data

Variable Name	Description
employment_status	category of employment
jacket_number	ID number
race	person's racial identity
sex	male or female
citizenship	person's country
age_at_arrest	person's age
days_in_jail	whole number greater than 0
offense_level	category of offense
high_level_crime	category of crime
city	of person's residence
state	of person's residence

CU Jail Data

What questions might a data analyst have about this dataset?

What might be their goals for analysis?

Example: CU Jail Data

Goal: We are interested in the proportion of the races of folks in the Champaign or Urbana.

We could use a bar plot to visualize that.

```
data cujail;  
  length employment_status $25. race $12. sex $6.  
    citizenship $8. offense_level $24. high_level_crime $12.  
    city $24. state $8.;  
infile 'location/ccso_subset_2.csv' dlm=',,' dsd missover  
  firstobs=2;  
input employment_status $ jacket_number race $ sex $  
  citizenship $ age_at_arrest days_in_jail offense_level $  
  high_level_crime $ city $ state $;  
run;  
proc sgplot data=cujail pctllevel=graph;  
  vbar city /group=race groupdisplay=cluster ;  
  where city in ('CHAMPAIGN','URBANA');  
run;
```

Example: CU Jail Data

How might we tweak or enhance this plot?

What alternative plots might accomplish the goal?

Exercise: CU Jail Data

1. How many people who were arrested lived outside of Champaign-Urbana?
2. When focusing on age at arrest, how many observations are missing?
3. Has anyone over the age of 70 and in Urbana been arrested?
4. Create a scatter plot of days in jail vs age at arrest for nonmissing ages only. What do you see?
5. Now create a scatter plot of days in jail vs age at arrest for black and white people arrested ages 35 and older. Describe the plot.