Adult Dataset

Description:

This Data is conducted from 1984 until 2010. The samples were veterans of the US Air Force and made by Ronny Kohavi and Barry Becker to analyze the adult financial status for further social studies. The extraction was done by Barry Becker from the 1984 Census database. The prediction task is to determine whether a person's income is over 50K a year.

Variables:

Income: >50K, <=50K. (0 is >50K and 1 is <=50K)
age: continuous.
workclass: a term to describe a person’s employment status
fnlwgt: financial weight.
education: An individual's highest education degree
education-num: The highest education degree of an individual in numeric form.
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
Occupation: the types of occupation
relationship: the individual’s relationship to others.
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
sex: Female, Male.
capital-gain: continuous.
capital-loss: continuous.
hours-per-week: working hours per week.
native-country: the individual’s original country.

Code to Import

Method 1: Download the CSV from the file directly

-To be more specific, students can download the CSV file from Canvas directly to

their own PC's desktop and then open an Rmd file and store it on the desktop as well. After doing this, students can apply this code to open the data set directly.

```{r}
read.csv("adult1.csv")->adult1
head(adult1)
```

Method 2: Download the raw dataset from the open data website and modify it for better use

-This adult data set can be download from the open data website: https://archive.ics.uci.edu/ml/datasets/adult.
-We can download the data through the hyperlink and then input column names for better visualization.
#this dataset's column names are not defined and we can put the column names according to the website's description.

```{r}
adult1<-read.csv(url("https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data"), header = F)
colnames(adult1)<-c("age","workclass",
                    "fnlwgt","eduaction",
                    "education_num",
                    "marital_status",
                    "occupation",
                    "relationship",
                    "race","sex",
                    "capital_gain",
                    "capital_loss",
                    "hours_per_week",
                    "native_country",
                    "income")    #fct variable
head(adult1)
```