Doris Chen

1. Using Pandas

time: 11am 1pm 1pm 2pm 4pm

temperature: cold warm warm hot hot

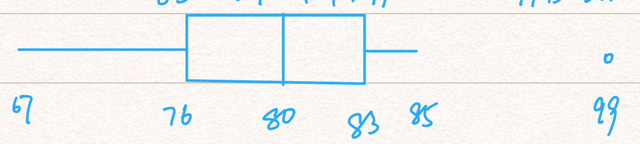
	price	gender	local time	temperature
median:	1.0	N/A	01-01-01 01:00pm	warm
mode:	1.0	F	01-01-01 01:00pm	hot, warm
mean:	2.0	N/A	using pandas 01-01-01 01:24pm	N/A

2. Median = 80 $Q_3 = 83$ $Q_1 = 76$ $IQR = Q_3 - Q_1 = 7$

check outlier: $max = 83 + 1.5 \cdot 7 = 93.5 < 99$

$min = 76 - 1.5 \cdot 7 = 65.5$

$83 + 3 \cdot 7 = 104 > 99$ $\therefore$ 99 is an outlier



Video

a) Because it's difficult to know what audiences really interested in, the lecturer cannot do live coding or dig deeper in how it works.

b) scrap: getting hold of the data we need
produce reliable dirty data
get data off web (web APIs)

clean : drudge-work of cleaning data
easy to locate duplicate records, fix dodgy date-strings, find

missing fields

Explore: Explore stories with annomalies hidden in the data

Result of serch can be saved

Deliver: Deliver data from database with a few lines of code

Transform: Selected reflections of the dataset are presented allow use to explore them interactively

c) They are interested in the trends and scenes forusing.

stories

-1