Lecture #05
Exploring Data with Pandas

Mar. 2020

Real Python

Introduction to Pandas

Exploring Data with Pandas

Data Cleaning Basics

Data Aggregation

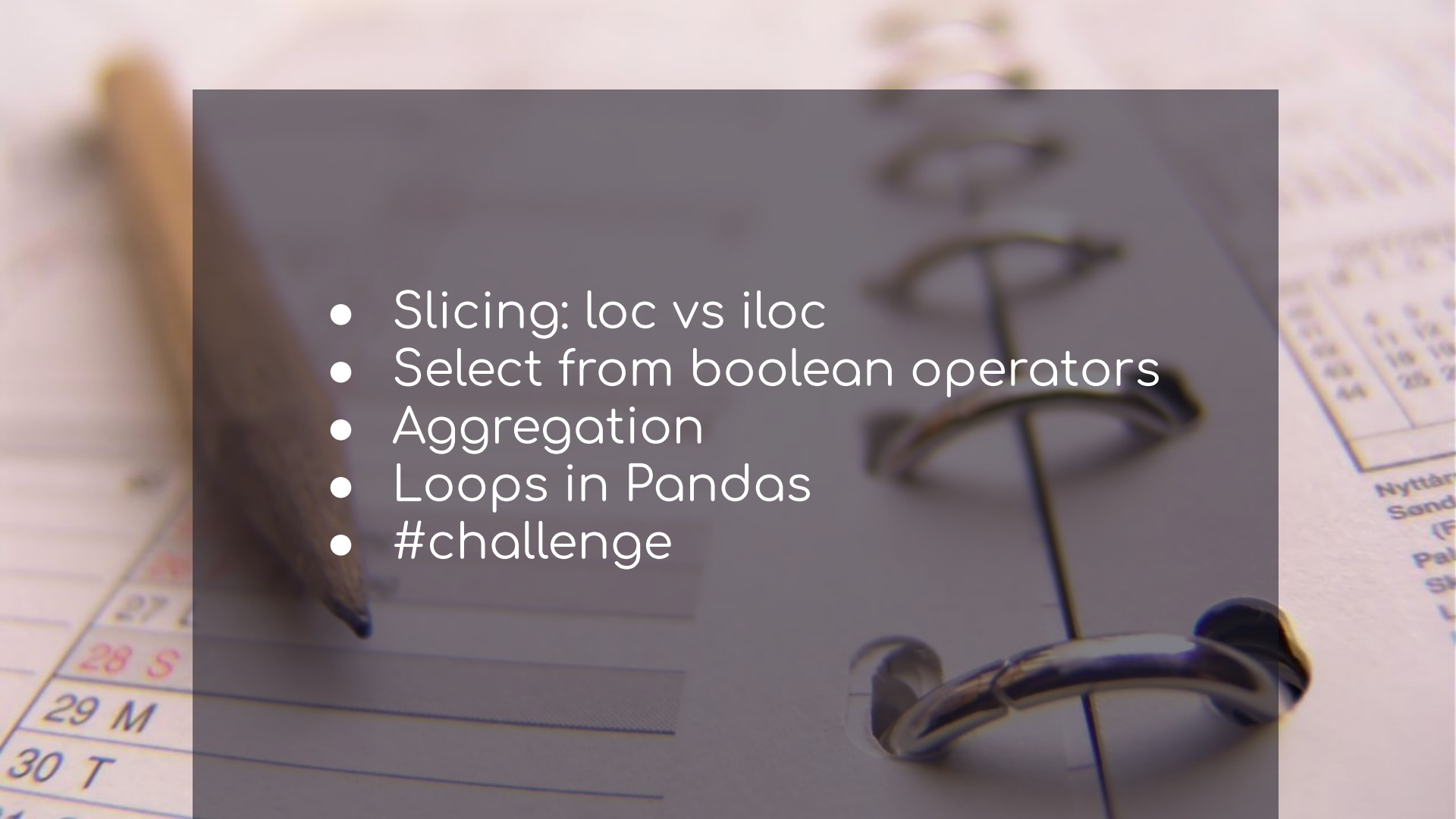<<Project 01>>

Combining Data with Pandas

Transforming Data with Pandas

Working with String in Pandas

Regular Expression

Working with missing and duplicate data

<<Project 02>>

- Slicing: loc vs iloc
- Select from boolean operators
- Aggregation
- Loops in Pandas
- #challenge

Update from repository

git clone https://github.com/ivanovitchm/datascience2020

Or ....

git pull

GitHub

iloc vs loc

df.iloc[1]

df.iloc[1]

df.loc[1]

df.loc[1]

1. Let's say that we wanted to select just the Brazilian companies from the Global 500
2. What companies are headquartered in California?
3. List the 05 companies with the largest number of employees? What is the total number of employees of these companies?
4. We want to find the companies more than 200 billion in revenue that are headquartered in China.
5. What the average revenue was for each country in the data set?

#challenge

https://plot.ly/python/plotly-express/

Plotly | Open Source Graphing Libraries | Python | Plotly Fundamentals | Plotly Express

◆ Edit this page on GitHub   🚀 launch binder

Navigation

Plotly Express

A single import with built-in datasets

Scatter and Line plots

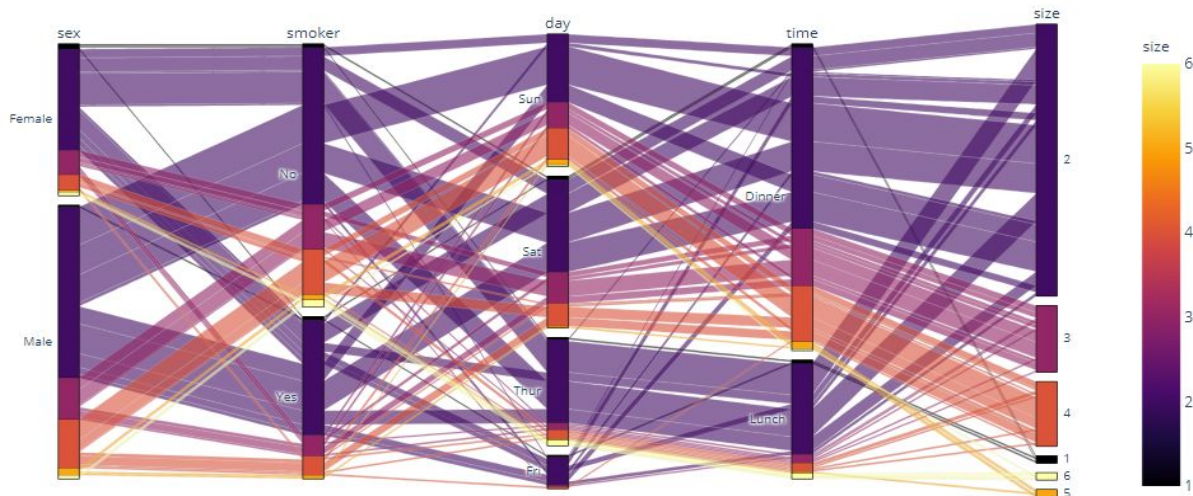Visualize Distributions

Ternary Coordinates

Images

3D Coordinates

Polar Coordinates

Maps

← Back To Python

```python
import plotly.express as px
df = px.data.tips()
fig = px.parallel_categories(df, color="size", color_continuous_scale=px.colors.sequential.Inferno)
fig.show()
```

Data Science Interview Questions

<<Optional Non-Graded>>

1. Given a sorted array of integers, how can you find the location of a particular integer x?
2. How can you quickly computer 2^x ?
3. What is the value of 15^18^15 ?