

Programação Funcional

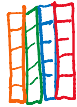
↳ pandas (A base, display, data.table)
↳ (spark) → tempo

arrumar dados: organizar, arrumar, corrigir, limpar 60-90%
→ tempo

analisar dados: análise, resumos, processar 40-50%
→ tempo



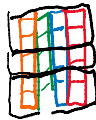
Quadro de dados
Base " "
Planilha " "



importar



ordenar (sort)



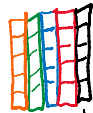
grupos (groupby)



last, first, head, tail



agg, apply → agregação → aplicar
(tabelas dinâmicas e/o excel, mas muito + flexíveis)
funções do pandas, da numpy e próprias



criar colunas



filtrar linhas



selecionar colunas

data frame

df.loc [l,c]
linhas
colunas

filtrar e modificar grupos de dados com o método loc

df.at [l,c]
linha
coluna

filtrar e modificar elementos individuais com o método at

1. Obter o volume médio por talhão
↳ desvio padrão

Obs: filtrar apenas as últimas medições de cada parcela.
↳ mais recente

- ① Importar os dados `pd.read_excel`
- ② Ordenar os dados e obter last `sort_values last`
- ③ Agrupar e agregar `agg combinada mean std`

2. Obter a ordem de colheita para o próximo ano
↳ identificar quantos km a máquina vai percorrer por conta própria e no caminho prancha.

- ④ Filtrar idade > 7.5
- ⑤ Ordenar por volume médio
- ⑥ Computar de → para com shift
- ⑦ Computar a distância entre linhas sequenciais com shift
- ⑧ Criar coluna prancha
- ⑨ Modificar parte da dataframe com `loc [l, c]`
- ⑩ Agrupar por prancha, selecionar coluna dist, somar dist por grupo