

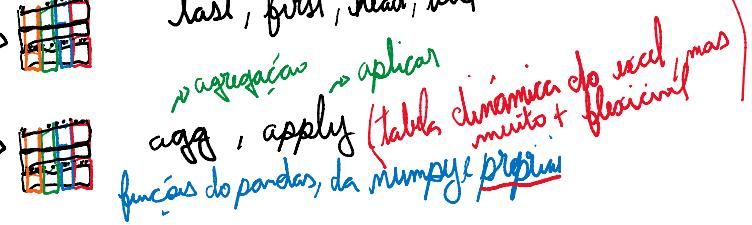
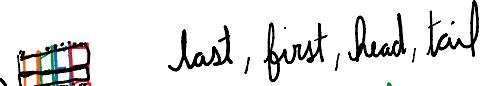
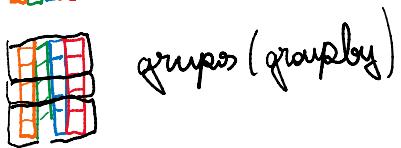
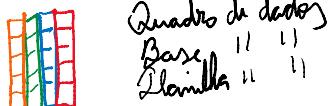
Programação Funcional

↳ **pandas** (R base, dplyr, data.table)

↳ (Spark) tempo

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

analisar dados: análise, resumos, processar 90-10%



funções do pandas, da numpy e próprias



Selecionar colunas

df.loc [l, c]
linhas
colunas

filtrar e modificar
grupos de dados com
o método loc

data frame

df.at [l, c]
linha
coluna

filtrar e modificar
elementos individuais
com o método at

1. Obter o volume médio por faixas

↳ 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 agrregar agg combinação mean std

2. Obter a ordem de colheita para o próximo ano

↳ identificar quantos km a máquina vai percorrer
por conta própria e no caminhão/prancha.

④ Filtrar idade > 7.5

⑤ Ordenar por volume médio

Computar di → para com shift

Computar a distância entre linhas subsequentes com shift

Gravar coluna prancha

Modificar parte da data frame com loc [l, c]

⑪ Agrupar por prancha, selecionar coluna dist, somar dist por grupo