



ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO
Licenciatura em Engenharia Informática

SISTEMAS DE INFORMAÇÃO

EXPLORAÇÃO E DESIGUALDADE LABORAL GLOBAL ATRAVÉS DE DADOS ABERTOS

Engenharia de Dados

João Augusto Costa Branco Marado Torres



Beja, dezembro de 2025

INSTITUTO POLITÉCNICO DE BEJA
ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO
Licenciatura em Engenharia Informática
SISTEMAS DE INFORMAÇÃO

EXPLORAÇÃO E DESIGUALDADE LABORAL GLOBAL ATRAVÉS DE DADOS ABERTOS

Engenharia de Dados

João Augusto Costa Branco Marado Torres

Trabalho realizado no âmbito da unidade curricular de Sistemas de Informação

ORIENTAÇÃO

Dr.ª Isabel Sofia Sousa Brito

Beja, dezembro de 2025

Júri

Responsável: Dr.^a Isabel Sofia Sousa Brito

Vogal: Dr. João Paulo Trindade

Vogal: Dr.^a Elsa da Piedade Chinita Soares Rodrigues

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

ETL *Extract, Transform, Load*

FAIR *Findable, Accessible, Interoperable, Reusable*

FLOSS *Free Libre and Open Source Software*

FMI Fundo Monetário Internacional

GUI *Graphical User Interface*

IED Investimento Estrangeiro Direto

ILO *International Labour Organization*

OLAP *Online Analytical Processing*

ONU Organização das Nações Unidas

PIB Produto Interno Bruto

POSIX *Portable Operating System Interface*

PPC Paridade do Poder de Compra

URI *Uniform Resource Identifier*

1 Introdução

As transformações recentes do capitalismo por conta da crise da ordem têm sido acompanhadas por um agravamento das desigualdades económicas e sociais, particularmente no domínio do trabalho e da distribuição do rendimento à escala mundial, levando à austeridade e reformas laborais. Diferenças persistentes nos níveis salariais e na estabilidade do emprego evidenciam assimetrias estruturais entre países e regiões, onde pessoas vivem bem numa parte do mundo em troca de outras viverem na miséria.

Este é o tema do trabalho: exploração global do trabalho e a desigualdade num nível global.

A disponibilização de *datasets* por múltiplos organismos internacionais e instituições públicas cria possibilidades para a análise destas desigualdades. Em particular, dados ao longo do tempo (*time series*), usando engenharia de dados e análise, permitem observar a evolução histórica de indicadores laborais e identificar tendências de longo prazo. Exemplos de tais indicadores podem ser as participações do rendimento do trabalho no PIB Produto Interno Bruto (PIB), coeficiente de Gini... (ver Apêndice ?? para definições e fórmulas). Era interessante conseguir analisar como a divisão do trabalho na dinâmica imperialista e neoliberal atual se manifestam nas acumulações desiguais, nas dependências imperialistas, na precariedade, nos salários, nas condições de trabalho...

Apesar da abundância de dados estatísticos disponíveis, a análise da desigualdade laboral global permanece fragmentada. Os dados encontram-se dispersos por múltiplas fontes e são frequentemente analisados de forma isolada, através de indicadores únicos.

Esta separação dificulta a compreensão das relações entre salários, rendimento, condições de trabalho e ação direta laboral (como as greves), bem como a identificação de contradições estruturais do capitalismo.

1.1 Objetivos

O objetivo geral deste projeto é investigar a relação entre a produtividade e parcelas salariais. Avaliar a dependência das trocas comerciais e do Investimento Estrangeiro Direto (IED) no *labor share*. Analisar as principais razões para as greves, por exemplo, se o salário não sobe por muito tempo. É apresentar as contradições do capitalismo usando dados *timeseries* abertos num formato multidimensional e ferramentas livres demonstrando a viabilidade técnica e ética dos mesmos. Pretende-se construir um modelo de dados que permita observar, de forma conjunta, diferentes indicadores relacionados com salários, rendimento e condições laborais, entre países ou regiões e ao longo do tempo.

De forma mais específica, o projeto visa:

- recolher e integrar dados provenientes de múltiplas fontes abertas;
- estruturar esses dados num modelo multidimensional adequado à análise;
- produzir análises e visualizações que evidenciem padrões, assimetrias e tendências relevantes.

1.1.1 Questões de investigação

Existe uma relação sistemática entre produtividade do trabalho e *labor share*?

Essa relação difere entre países centrais e periféricos?

Indicadores de dependência externa (IED) estão associados a menor *labor share*?

Existe associação entre precariedade e ação direta?

1.1.1.1 Hipóteses

Em contextos de aumento da produtividade do trabalho, a parcela salarial do PIB tende a permanecer estagnada ou a diminuir.

A inserção subordinada na economia global está associada à transferência sistemática de valor do trabalho de países periféricos para países centrais.

1.2 Abordagem e Estrutura do relatório

Para alcançar estes objetivos, o trabalho adota uma abordagem baseada em engenharia de dados e o modelo de análise *Online Analytical Processing* (OLAP). O processo inclui a aplicação do padrão *Extract, Transform, Load* (ETL), a construção de um *data warehouse* e a realização de análises multidimensionais via "cubos" (uma base de dados multidimensional em conceito), tabelas e visualizações.

O relatório encontra-se organizado em três capítulos principais. Após esta introdução, o capítulo de desenvolvimento apresenta o enquadramento teórico, a metodologia adotada, a análise dos dados e a discussão dos resultados. Por fim, o capítulo de conclusão sintetiza os principais resultados, discute as limitações do estudo e aponta possíveis direções para trabalhos futuros.

2 Desenvolvimento

2.1 Enquadramento teórico

Enquadramento metodológico geral

2.1.1 Economia política

2.1.2 Desigualdade global

2.1.3 Divisão internacional do trabalho

2.1.4 Limitações dos dados estatísticos

Com informações como estas, como se mede exatamente a exploração, a dependência do imperialismo, a precariedade, ou a extração de valor. Não é bem o objetivo medir a exploração diretamente de dados monetários como o PIB e os outros. Talvez isso nem seja possível e não é só uma questão de eu não o saber fazer. O meu objetivo é pegar nesses dados e usá-los para fazer uma análise empírica, quer dizer, observando os dados, e olhando para o mundo em que vivemos, e chegar à conclusão que a minha análise bate certo com a realidade, e é consistente com teorias da economia política. Eu penso que os dados que escolho usar têm uma relação suficientemente forte para conseguir testar a minhas hipóteses, apesar das limitações. Seria parecido às formas de como é medida a qualidade de vida, que não é algo que se meça diretamente.

Outra coisa é que pelos dados serem de fontes diferentes, podem usar definições e metodologias diferentes para calcular o mesmo (teoricamente), que depois pode causar-me problemas se eu não tiver cuidado. É algo com que preciso de ter atenção na etapa ?? do ETL.

2.2 Metodologia

2.2.1 Escolha das técnicas e *software*, e técnicas de análise

O meu foco é em criar algo que é facilmente reproduzível, sem ter que te dar indicações de onde tens que clicar na interface gráfica para elaborar uma ação específica. Por isso, decidi usar R ou Python para o ETL, com ajuda de *scripts Portable Operating System Interface* (POSIX) e *Makefiles*.

No início, ainda pensei em usar um simples *script* POSIX. Mas depois achei que ia se transformar em algo muito grande, complicado, e com código repetido. Passei logo para a ideia de usar Makefiles. Eles vão-me permitir orquestrar todos os passos a executar neste trabalho, e caso eu mude algo no *transform*, por exemplo, consigo pedir para executar tudo, mas na realidade ele vai dar passar o passo da extração já que ele não foi afetado por nenhuma mudança, então poupo um pouco de tempo.

Para o *load*, usamos a base de dados DuckDB, que na realidade, ela permite fazer o ETL inteiro através do que ela oferece fora da caixa, e se necessário, extensões. Por exemplo, conseguimos criar tabelas a partir de documentos usando o protocolo HTTP, por exemplo, usar o *statement* UNPIVOT se necessário, e no final, usar a operação de conjuntos UNION [ALL] BY NAME numa *query*. Mas nós apenas a vamos usar pelo facto de ela ter sido feita para conseguir fazer operações OLAP de raiz, rapidamente. Com certeza uma experiência para os tempos livres.

```
1 CREATE OR REPLACE TABLE flights AS
2   FROM 'https://duckdb.org/data/flights.csv';
3
4 UNPIVOT monthly_sales
5 ON COLUMNS(* EXCLUDE (empid, dept))
6 INTO
7   NAME month
8   VALUE sales;
9
10 SELECT * FROM capitals
11 UNION BY NAME
12 SELECT * FROM weather;
```

Para a parte de análise, Jupyter Notebooks ou R Markdown, ou apenas Python ou R com bibliotecas como Gnuplot ou parecido. Depois consigo gerar PDFs e PostScripts.

No final, se tiver tempo, experimentar codificar os dados para um formato de 4 estrelas no mínimo, RDF ou JSON-LD.

Penso em usar os serviços da Wikimedia Foundation também, caso precise de informação extra, ou se eu quiser contribuir para o projeto com a informação que conseguir retirar da análise.

Durante a análise, vou experimentar várias formas de mostrar as informações: tabelas a partir do *pivot*, correlações, regressões...

Quero depois disponibilizar os dados tratados em formatos abertos, usando *Uniform Resource Identifier* (URI) sempre que possível para identificadores de países ou anos. Garantir que está tudo FAIR. E documentar neste relatório o processo ETL.

Estas decisões não foram feitas ao calhas. Eu tenho uma ideia muito forte de como todos devemos usar e criar software, incluindo, mas especialmente para trabalhos académicos. A explicação para isso está no Apêndice ??.

2.2.2 Estrutura do projeto

Terá um ficheiro `README.md` que servirá como uma introdução para quem pela primeira vez olha para este projeto, sem nenhum contexto.

Na pasta `data` vi estar guardado tudo o que são dados. Teria uma pasta `raw` com os dados acabados de serem retirados da *source* e uma `cleaned` que guardasse as transformações dos anteriores com todas as mudanças necessárias para o seu uso na análise.

O que vai fazer as tarefas mencionadas em cima vão ser uma série de *scripts* localizados na pasta `cleaned` com executáveis.

O código para a criação do relatório e quase tudo relacionado à documentação estaria na pasta `docs`.

Como a minha ideia é usar Python ou R, faz sentido que eu acabe por usar Jupyter Notebooks ou R Markdown para complementar a documentação, isso estará na pasta `notebooks`.

Código SQL fica na pasta `sql`.

2.2.3 Fontes de dados

Durante a minha pesquisa achei múltiplas potenciais fontes de dados na Internet capazes de me ajudar nesta tarefa:

Organização Internacional do Trabalho Disponibiliza dados sobre empregabilidade, salários, *labor share*, sindicalização, horas de trabalho;

Banco Mundial Para dados como o PIB, Gini, abertura comercial, fluxos de capital;

Our World In Data Oferece diversas informações como a desigualdade, salário mínimo, ou sobre o trabalho informal e desemprego;

Wikidata Para qualquer metadados que eu necessite, sobre países, empresas, sindicatos e protestos relacionados ao trabalho;

Confederação Sindical Internacional Eles fazem os *"ITUC Global Rights Index"* cada ano onde eu consigo encontrar informações como violações ao direito dos trabalhadores;

Base de dados sobre Desigualdade Mundial Uma base de dados sobre a evolução histórica da distribuição mundial do rendimento e da riqueza, dentro como entre países.

Dado o caráter dos dados das fontes utilizadas, sem nenhuma relação inicial, a análise baseia-se em indicadores indiretos, entendidos como aproximações empíricas a processos sociais mais complexos, ou seja, vamos relacionar esses valores para termos uma aproximação dos valores que realmente queremos, e que na realidade, são impossíveis de medir: exploração. Como foi falado na ??

Por exemplo, o PIB representa o valor de riqueza produzido num espaço de tempo num determinado local. Com dados sobre o número de horas trabalhadas durante esse espaço de tempo, eu consigo ter a razão do PIB por hora de trabalho. Quanto maior esse valor, então mais riqueza foi gerada em uma hora, maior a produtividade. Depois temos dados como o *labor share*: qual é a percentagem do rendimento gerado num espaço de tempo num determinado local — pode ser a percentagem do PIB — que foi para quem gerou o rendimento (trabalhadores), por exemplo, em salários, ao invés de ir para o capital. Com isto, posso fazer uma hipótese empírica de que, em economias periféricas, aumentos de produtividade não se traduzem em aumentos proporcionais da parcela salarial.

2.2.4 *Extract, Transform, and Load*

2.2.4.1 *Extract*

Como pôde ser visto em ??, eu mencionei várias fontes de dados, mas pensei que usar todas elas de uma só vez iria complicar-me a vida muito. Então, para começar, escolhi usar apenas duas: a Organização Internacional do Trabalho — vou começar a mencionar por International Labour Organization (ILO) — e o Banco Mundial para começar. A extração tem que ser feita de modo que no futuro seja simples adicionar outras fontes de dados.

Para o ILO, decidi extrair informações como o *labour income share as a percent of GDP (%)*, a quantidade de contratos de emprego (formais e informais) divididos por sexo, atividade económica e setor (público ou privado) do trabalhador, e a média do salário por hora também dividido por sexo e atividade económica do trabalhador. Do Banco Mundial, de cada país, extraí o PIB (no valor atual do dólar dos Estados Unidos da América), o *Gini index*, e as entradas líquidas de IED (registado na balança de pagamentos e expresso em dólares).

Comecei pelo ILO. Descobri que eles oferecem uma biblioteca R — (Bescond et al., 2025) — para facilitar o meu trabalho. Ficou decidido logo aqui que eu iria usar R ao invés de Python (apesar de que eu já estava inclinado para usar o R). Aprendi o básico de R a partir do (R Core Team, 2025). Depois, vi como podia usar a biblioteca que encontrei, e espreitei as bibliotecas listadas nos *imports* e *suggests* desta para saber o que faziam, já que poderiam dar-me jeito — e deram. Qualquer dúvida, recorri ao («Perplexity AI», s.d.).

Quero mencionar que bem no início era para eu usar o (OpenRefine Community, 2025), por ser o que me parecia o mais semelhante ao Power Query. Ele permite o *extract* e *transform* dos dados, e essas ações, ficam guardadas num histórico, num ficheiro JSON, que serviria para reproduzir as operações entre computadores. Mas isso requisita usar a *Graphical User Interface* (GUI) mesmo assim; já não achei reproduzível o suficiente. Entretanto, passei também os olhos na possibilidade de usar (Apache Software Foundation, s.d.),

(Spotify AB, s.d.) e (Prefect Technologies, s.d.). Ficam aqui as menções.

Criei o ficheiro `extract.R` que apenas fazia o *download* dos *datasets*.

Passando para o Banco Mundial, achei que seria mais simples apenas usar (Nikšić, s.d.). Renomeei o ficheiro anterior para `extract_ilo` e criei o `extract_worldbank.sh`. Descobri logo que o *link* que eles disponibilizavam na API para fazer o *download* do *dataset* em *comma separated values* na realidade era um arquivo comprimido, mas lá dei a volta a esse problema. Com o ficheiro com os dados, vem um com metadados para cada código de país, e um com uma descrição sobre a informação que eu tinha carregado da Internet.

Comecei a criar o `Makefile`. Criei uma regra que executa todos os ficheiros na pasta `bin` que comecem com `extract_`. Para não ter que estar sempre a editar o ficheiro e os `targets` com uma mudança na extração, usei a criação de um ficheiro sem nada escrito, mas que a presença dele indica que o E do ETL já foi concluído, e que depois pode ser usado para dependências.

2.2.4.2 Transform

Usei o mesmo formato de nomear os ficheiros, e a mesma ideia no `Makefile`, mas agora a regra `transform` depende do `extract`.

Usei algumas das sugestões de bibliotecas que encontrei anteriormente, nomeadamente o (Wickham et al., 2023) para facilitar a manipulação dos *data frames* e (Wickham & Posit Software, 2025) para manipular *strings*. Também encontrei o (Wickham et al., 2025) que me deixa usar colunas dos *data frames* como variáveis, facilitando o processo todo.

Nos dados do ILO queria renomear as colunas em todos os ficheiros para o mesmo, e remover a linha que podia ser calculada a partir das outras (o total do setor económico).

Para o Banco Mundial, quis renomear os ficheiros para algo menor que explicasse o que estava presente neles. Tive que fazer um *unpivot* criando a coluna `year`. Mudei o nome de algumas colunas também para ficar igual ao do ILO e juntei alguns dos metadados que tinham vindo juntos.

Foi inicialmente adotada uma estratégia de consolidação dos indicadores num único conjunto de dados antes do carregamento. Criei um único ficheiro — `integrate.R` — onde eu leio cada ficheiro de dados, e uso o código do país e no ano de cada linha para criar linhas que agregam todos os valores correspondentes a esses dois que os ficheiros separadamente possuíam. No *data frame* do salário por hora, tive que criar 3 novas colunas via *pivot*, um para cada tipo de representação do salário, e depois escolhi usar apenas o método Paridade do Poder de Compra (PPC), que me pareceu a melhor opção comparado com o valor em dólar ou em moeda social.

No entanto, esta abordagem revelou-se inadequada do ponto de vista de modelação multidimensional. Uma tabela não normalizada não é capaz de se tornar numa *data warehouse*. Estaríamos a adicionar uma etapa a mais ao processo ETL que não existem e é conceitualmente errado. Não existem tabelas de factos, nem de dimensões. Não existem chaves primárias nem estrangeiras. E também não existe um *star/snowflake schema*.

Desnormalização e agregação são tarefas que só vão ser necessárias mais tarde durante a análise.

Sem uma *data warehouse*, não temos modelo multidimensional, não conseguimos fazer OLAP, e torna-se complicado fazer análises.

A tarefa aqui no *transform* é preparar os dados para a criação da *data warehouse* na próxima etapa *load*. Então para isso já temos que ter um modelo multidimensional em mente, o qual foi o que faltou anteriormente, e por isso criamos uma etapa que estraga o propósito do trabalho. O nosso modelo multidimensional é detalhadamente explicado na ??.

Sempre a atualizar o `Makefile`.

Neste momento existiam muitas linhas cheias de `NA`, exatamente porque essa informação não existe. O que fazer? Por serem tudo números, uma das opções preencher todos os espaços vazios com 0. Mas isso não faz sentido.

- PIB 0 significava que nenhum dinheiro foi gerado, impossível;
- Coeficiente Gini 0 era dizer que não existiam desigualdades, impossível no capitalismo;
- Participação do rendimento do trabalho no PIB de 0 é algum tipo de escravidão;
- Média do salário por hora ser 0 também é exploração total do trabalhador.

IED 0 até é possível existir, mas improvável — o fluxo de entrada e saída tinha que ser o mesmo, anulando-se —, e a ausência do valor não significa que ele não existiu, apenas que não foi calculado, ou sei lá.

Devo remover as linhas que têm pelo menos uma falta de valor então? Não, porque isso normalmente acontece nos países do sul global, e isso tiraria a sua representatividade no *dataset*.

Existem muitas formas de tratar os valores em falta. Se falta dados em anos que ficam entre outros anos onde temos dados, podíamos interpolar esses valores com ajuda de *machine learning*. Neste caso, apenas deixamos as linhas da forma que elas estão, sem apagar nada também, por forma a preservar representatividade dos países periféricos. Na altura da análise, é decidido o que fazer em cada ocasião com os dados em falta.

2.2.4.3 Load

A decisão ficou entre («DuckDB», 2025) e («TimescaleDB», 2025) para PostgreSQL, que me lembrei de ambas as menções de vídeos separados do canal de YouTube (Fireship, 2025).

Optei pelo DuckDB por ser mais minimalista, estilo SQLite, enquanto para o outro eu ia precisar de um servidor PostgreSQL.

A cena do DuckDB é que foi desenhado para suportar OLAP tratando as tabelas não como um conjunto de linhas, mas mais como um conjunto de colunas, e por isso é rápido nas operações OLAP.

Para esta parte do processo ETL, apenas escrevi SQL que o DuckDB entende, e o `script load.sh` manda executar esse código no DuckDB e cria um ficheiro de base de dados persistente.

2.2.5 Modelo multidimensional

O modelo multidimensional proposto organiza os dados em torno de uma tabela de factos centrada em indicadores económicos e laborais, articulada com dimensões temporais, geográficas e estruturais.

Os indicadores do Banco Mundial foram modelados através de uma tabela de factos única. Cada valor para cada um daqueles indicadores representa esse valor numa área do nosso planeta, num determinado ano. A granularidade destes factos então é o país (em modo geral), o ano, e o indicador, permitindo a análise conjunta e comparativa de múltiplos indicadores macroeconómicos ao longo do tempo, evitando a mistura de métricas com diferentes unidades e propriedades de agregação. A tabela de factos observa os indicadores macroeconómicos para as suas dimensões: país observado; indicador (o que estamos a medir); e a data (neste caso o ano).

A tabela de dimensão país, vai ter as colunas (excluindo a chave primária) para: o código ISO; o nome do país; a região; e o grupo de rendimentos.

Depois pode ter outros conforme a necessidade.

O grupo de rendimentos, não é sempre o mesmo conforme o tempo vai passando. Mas nós vamos usar o rótulo atual desse país (o que vem nos ficheiros retirados da Internet), e que depois conforme o tempo passa e essa informação é atualizada, nós atualizamos o rótulo desse país.

A tabela de dimensão indicador, vai ter as colunas para: o código do indicador; o nome; a origem da informação; e a unidade do valor.

Os dados do ILO não permite existir apenas uma tabela de factos porque a granularidade e a tabela de dimensões é diferente em alguns casos. Apresentam maior granularidade e heterogeneidade do que os do Banco Mundial, exigindo a modelação de múltiplas tabelas de factos com diferentes granularidades, mantendo dimensões conformadas sempre que possível.

No caso do *labor share*, dá para adicionar à tabela de factos para o Banco Mundial, já que a granularidade é a mesma.

Para a tabela de factos sobre os salários, usamos os dados o salário médio por hora. A granularidade usa o país e o ano — como os outros — mas para além disso, também o sexo, a atividade/setor económico, e a forma de representação monetária.

Por último temos informação sobre os contratos (ou falta deles) de trabalho, informal ou formal. Para ambas as situações, a granularidade é a mesma, com o país, mês, sexo, atividade/setor económico, e setor institucional (público, privado).

Também existe uma tabela de dimensão acerca da qualidade da informação, usada por todas as tabelas de facto com dados da ILO.

O modelo multidimensional resultante assume a forma de uma constelação de esquemas estrela, com múltiplas tabelas de factos correspondentes a diferentes processos estatísticos (indicadores macroeconómicos, salários, emprego), partilhando dimensões conformadas como país, tempo, atividade económica e fonte.

Acabamos com um esquema *galaxy* (ou será constelação ou floco de neve?).

O nosso *dataset* como está agora, é incompleto, no sentido que não serve para responder a todas as questões apresentadas na introdução ???. Mas o *data warehouse* permite com que o modelo seja expandido mais tarde (especialmente com as informações sobre sindicalização e ação direta).

2.2.5.1 Tabelas de factos e dimensões

A tabela de fatos é composta por indicadores económicos com relação ao trabalho. Teria como medidas o *labor share*, as *real wages*, o Gini, a produtividade, fluxos de IED, e dias de greve. As dimensões seriam o tempo (ano, década), país, região, setor, tipo de trabalho (formal ou informal).

Para o "cubo", usaria essas dimensões, como o tempo e a geografia — países e regiões, norte e sul global, classificações dos países conforme o Fundo Monetário Internacional (FMI) e a Organização das Nações Unidas (ONU) —, como forma de "divisão" de dados para análise.

2.3 Análise

Na minha pesquisa para encontrar a melhor substituição ao Power BI, achei o seguinte:

- Metabase;
- Apache Superset;
- Grafana;

Todos depois teriam o mesmo problema que achei com OpenRefine.

Decidi que esta seria a melhor altura para usar RMarkdown e bibliotecas como o *ggplot2* que também tinha sido recomendado. Mas como não tive tempo para aprender como usar RMarkdown, usei apenas um RScript.

No futuro pretendo experimentar o Streamlit.

2.3.1 OLAP

2.3.2 Resultados principais

A fase de análise inicial baseou-se num conjunto de dados desnormalizado; a reformulação subsequente implementou um *data warehouse* multidimensional que permitiu a análise OLAP e a agregação teoricamente fundamentada.

A análise apresentada abaixo põe isso mesmo em prática.

2.3.2.1 Evolução global do *labour share*

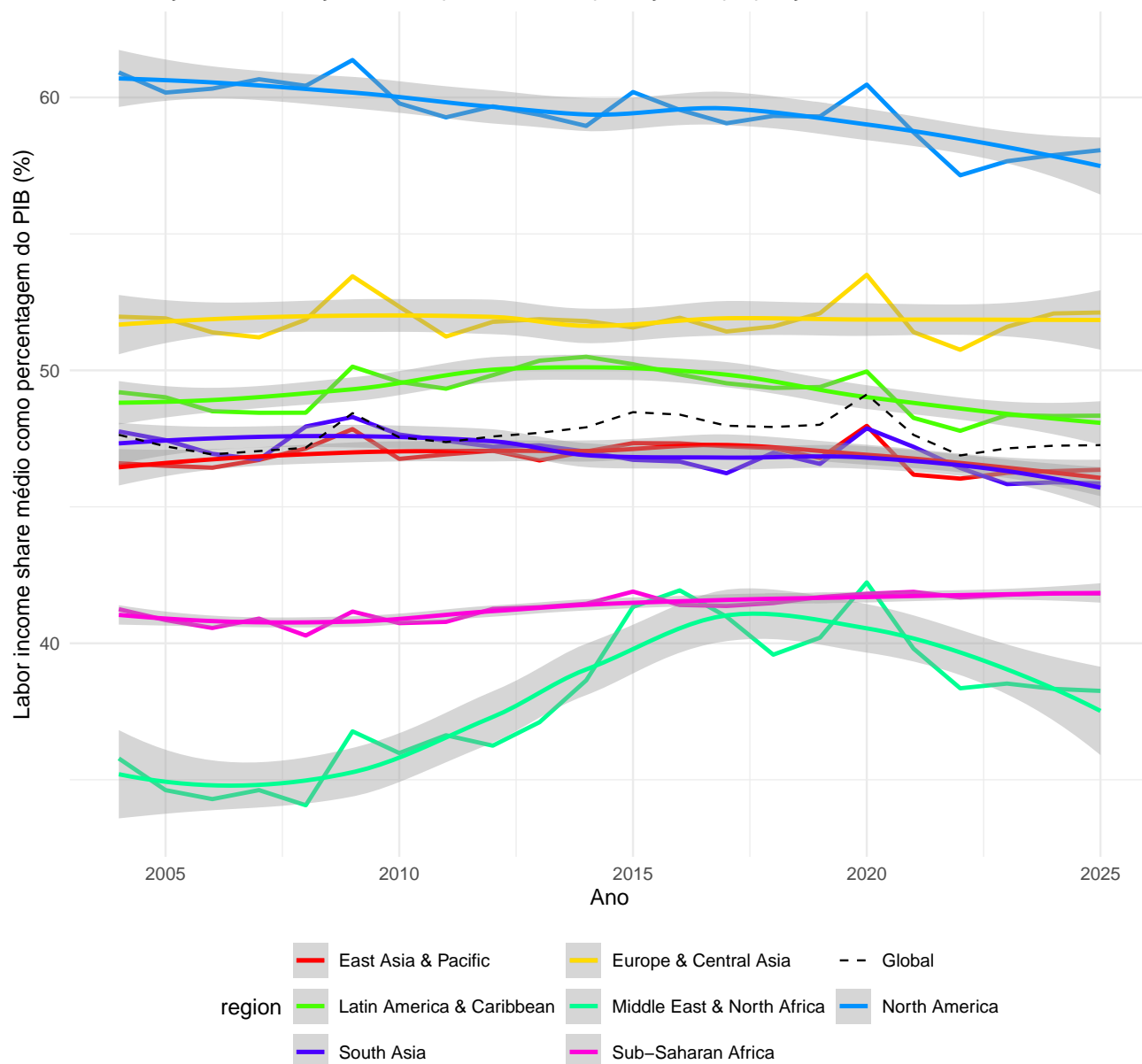
Tabela 2.1 *Visão geral do labour share por região (2004–2025, a cada 5 anos).*

year	East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	North America	South Asia	Sub-Saharan Africa
2005	46.50462	51.90579	49.01374	34.62096	60.1775	47.44467	40.85213
2010	46.75659	52.33425	49.57748	35.97861	59.7805	47.64883	40.73968
2015	47.33338	51.57227	50.22565	41.33578	60.1975	46.72233	41.89534
2020	47.96679	53.50062	49.95961	42.23030	60.4700	47.87667	41.81283
2025	46.35841	52.11966	48.34071	38.25619	58.0670	45.85650	41.82416

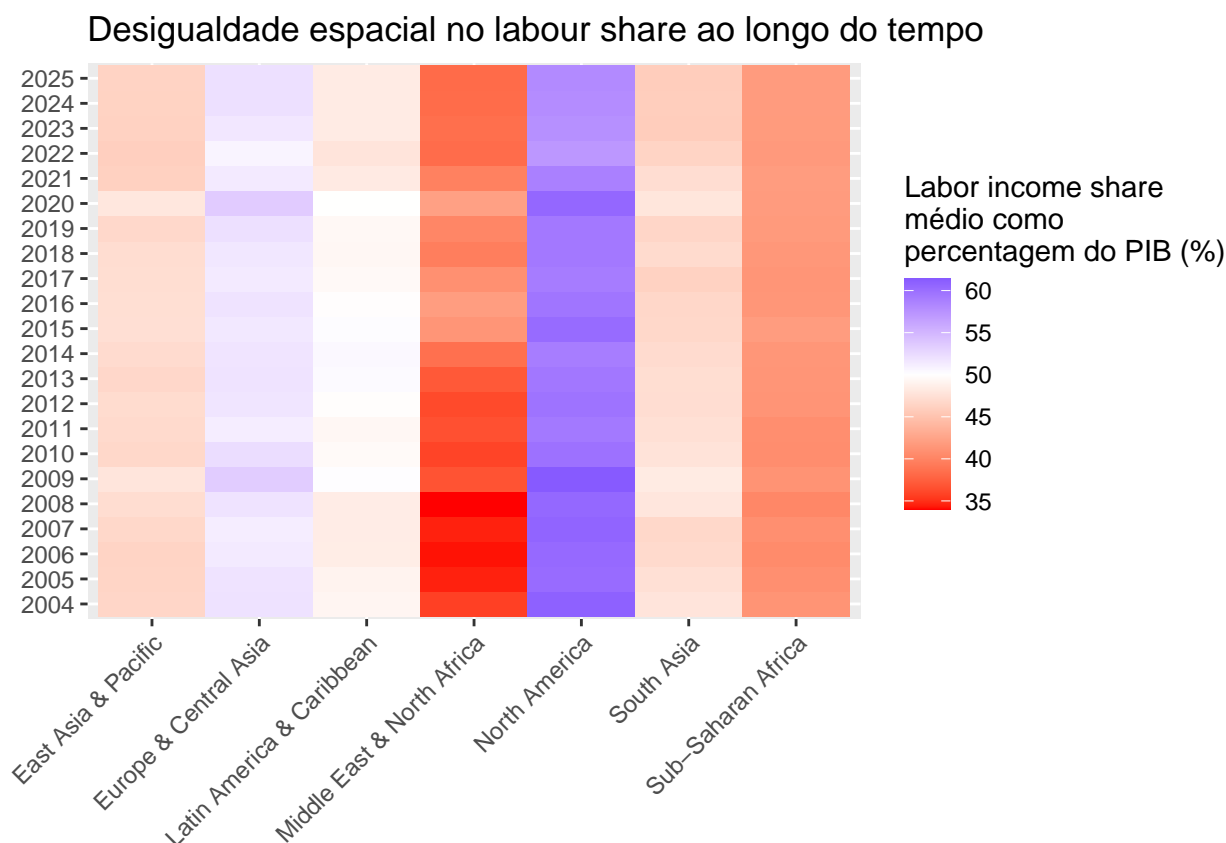
Conseguimos ver como os “países de velho mundo” são os únicos com um *labour share* acima dos 50%. A nível global, isso não se verifica. Embora que consigamos registar um declínio da parte da América do Norte de 61% em 2004, para 58% em 2025. A maioria das regiões mostra certa estabilidade no valor (o que não significa ser algo bom), tirando a América do Norte (com o declínio já mencionado), e o Médio Oriente e Norte da África. O Médio Oriente e Norte da África, Para além de terem o pior *labour share* registado em (quase) todos os anos (34%–42%), é bastante desregulado. Começou a subir perto de 2007–2008, quanto teve a crise global. São países com grandes riquezas petrolíferas, e que quando o preço do “ouro negro” subiu, o mesmo não se verificou na carteira dos trabalhadores. A Primavera Árabe ocorreu após essa crise, e é provável que, para acalmar as revoltas, os governos criaram reformas que aumentassem o *labour share* de certa forma. Também é possível notar o aumento a nível global no ano 2020, com o evento do início pandemia do SARS-CoV-2.

Tendências Globais do labour share(2004–2025)

Intensificação da contradição entre capital e trabalho: produção e apropriação de valor



Uma coisa é certa, com um *labour share* a nível global de cerca de 48%, e regiões onde o valor é 10% acima ou abaixo desse valor, com tendências descendentes nas regiões de ambos os casos. Existe claramente uma desigualdade regional, com os trabalhadores do norte global a ganhar uma maior parte da riqueza que a sua região gera. Apesar de declínios a nível regional, existe uma certa estabilidade global.



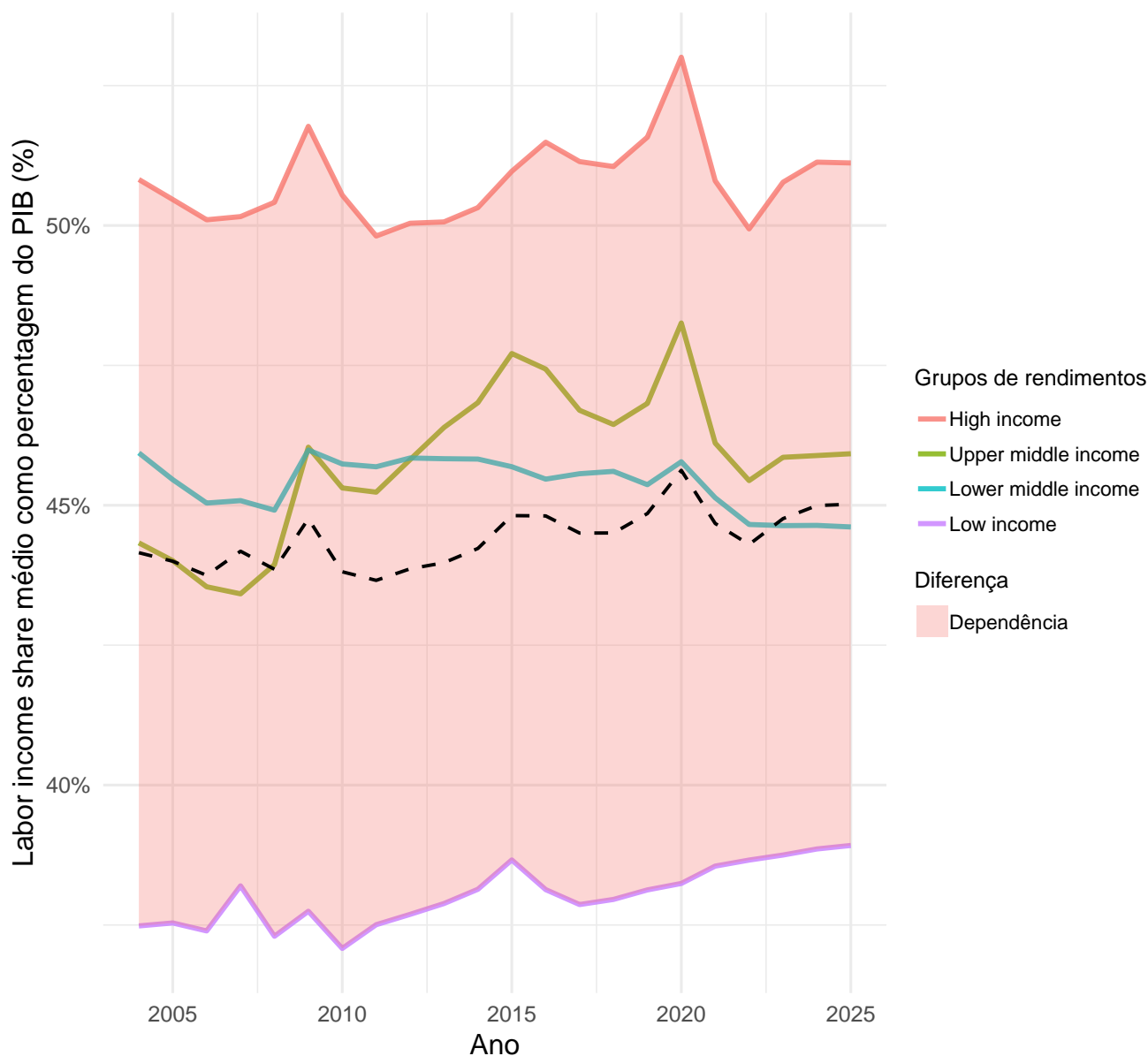
O *heatmap* claramente destaca os mínimos das regiões do Médio Oriente e Norte da África/Subsaariana em comparação com os máximos da América do Norte.

2.3.2.2 Comparação entre centro e periferia

Conseguimos ver claramente uma diferença enorme entre os países com maior rendimento e aqueles com menor rendimento. Existe uma diferença maior que 10% sobre o *labour share*. Os países de renda médio alta e de renda médio baixa estão quase na mesma situação, e acima da média dos extremos normalmente, com os países de renda média baixa a passar para baixo nos últimos anos. Conseguimos reparar que os de renda média baixa já tiveram em média maior *labour share* que os de renda médio alta até 2012. Mas também é de lembrar que talvez, os países que hoje estão categorizados como sendo de renda média alta, antes de 2010 era de renda média baixa, e vice-versa, como já foi notado anteriormente.

Divergência no labour share entre o centro e a periferia

Extração persistente de lucro da periferia para o centro



Uma possível forma de demonstrar a teoria da dependência mostrando diferenças persistentes no labour share entre economias centrais (alta renda) e periféricas (baixa renda), destacando a dinâmica da exploração.

Os países centrais de rendimentos altos mantêm cerca de 50% do labour share, enquanto os países periféricos de baixo rendimento estagnam em cerca de 38%, criando uma diferença estrutural de um pouco mais de 10% que visualiza o mecanismo de “dependência” e extração central da teoria da dependência.

2.3.2.3 Produtividade *versus* labour share

Agora, vamos tentar mostrar uma contradição entre estas duas coisas:

- Produtividade (produção por trabalhador) — medida em PIB por trabalhador (paridade do poder de

compra em dólares), reflete o aumento da capacidade técnica da mão de obra e os ganhos de eficiência dentro do capitalismo. Isto captura diretamente a produção por trabalhador, tornando-o ideal para revelar a contradição entre o aumento da produtividade e a estagnação do *labour share*. Como se concentra apenas nos trabalhadores empregados, evita distorções de fatores demográficos, como taxas de fertilidade ou reformados. O ajuste à paridade do poder de compra (PPC) garante a comparabilidade entre as diferentes regiões do globo;

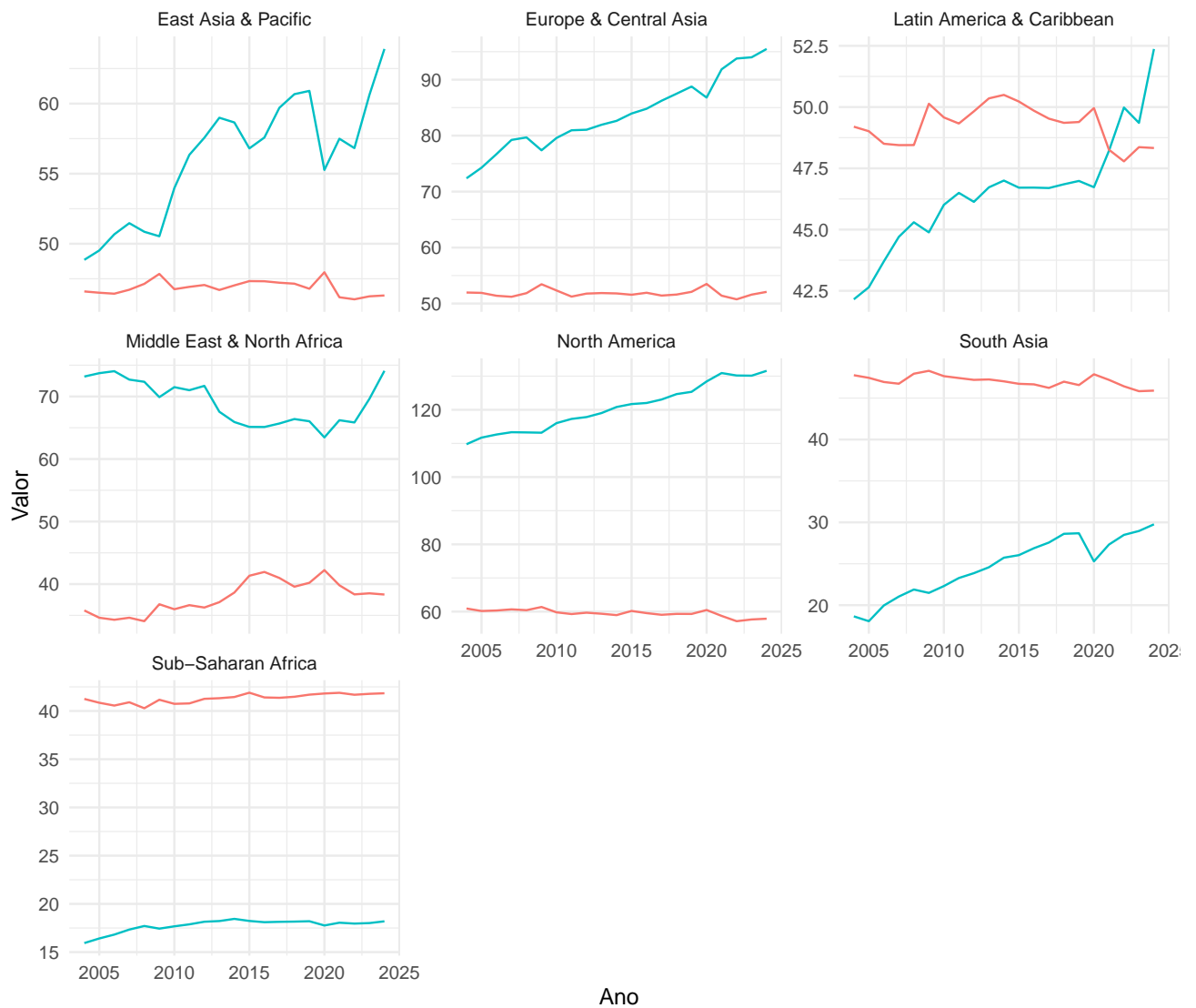
- Labor income share — que representa a parte da riqueza efetivamente capturada pelos trabalhadores.

A produtividade tende a aumentar devido à inovação, mas *labour share* muitas vezes estagna ou diminui — o que significa que o capital captura mais da mais-valia.

A análise seguinte deve ajudar a tornar explícito esta contradição: os trabalhadores aumentam cada vez mais a produtividade, enquanto a sua reivindicação coletiva sobre o lucro (representado pelo *labour share*) é mediada e limitada pelo controlo do capital sobre a produção.

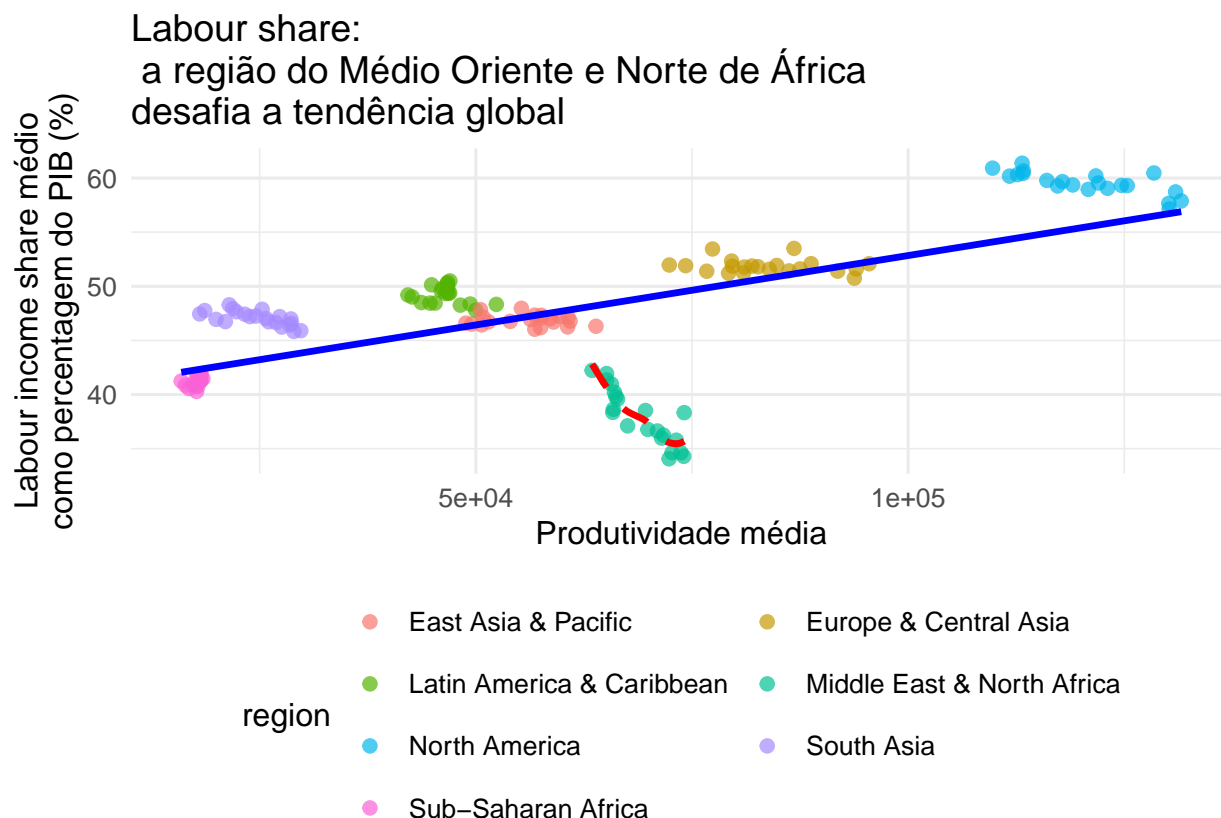
Aumento da produtividade versus estagnação do labour share

Visualizando a contradição central do capitalismo: produção coletiva, apropriação privada



— Labour income share médio como percentagem do PIB (%) — Produtividade x 1000 (PPC em dólares)

Na maioria das regiões (excluindo o Médio Oriente e o Norte de África até a altura do COVID), a produtividade só têm aumentado a nível global. Enquanto isso, temos o *labour share* basicamente estagnado.



Se a relação fosse negativa ou estável, tornava explícito: à medida que a produtividade cresce, a reivindicação dos trabalhadores não cresce.

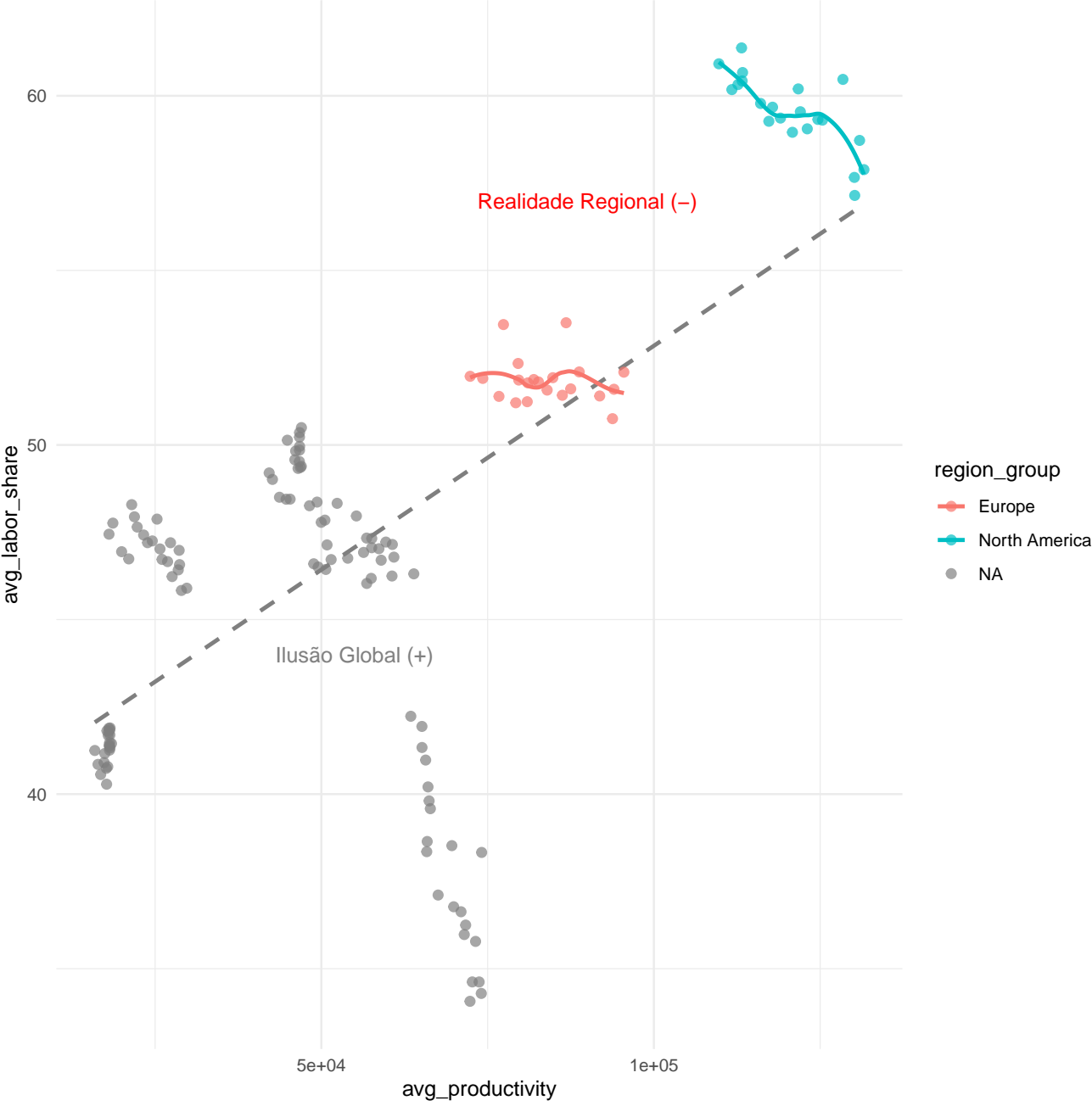
Porém, afinal vemos que os ganhos de produtividade traduzem-se parcialmente no *labour share*, **mas apenas a nível global**. E isso fica bem claro quando olhamos para a exceção: a região do Médio Oriente e Norte de África, onde os rendimentos do petróleo, por estarem sob controlo autoritário dos líderes da região, são eles que ficam com o mais da mais-valia. A produtividade têm aumentado também, porém os trabalhadores recebem cerca de 35%–40%, enquanto o capita acumula mais de 60%. O que isto mostra talvez na realidade é que esta é a única região em que os capitalistas não aumentam, nem que ligeiramente os salários, para manter os trabalhadores calmos, ao mesmo nível que do resto do mundo. São dois modelos de extração do lucro diferentes: a gerida e a descarada.

Tabela 2.2 *Visão geral do labour share por região (2004–2025, a cada 5 anos).*

region	slope	std.error	p.value	statistic
South Asia	-0.0001262	0.0000310	0.0006620	-4.0639675
Sub-Saharan Africa	0.0003493	0.0001454	0.0267061	2.4017787
East Asia & Pacific	-0.0000148	0.0000271	0.5912831	-0.5462030
Middle East & North Africa	-0.0006366	0.0000808	0.0000002	-7.8815888
North America	-0.0001202	0.0000219	0.0000264	-5.4987752
Latin America & Caribbean	-0.0000871	0.0000758	0.2652746	-1.1478594
Europe & Central Asia	-0.0000191	0.0000225	0.4047215	-0.8521926

region	slope	std.error	p.value	statistic
--------	-------	-----------	---------	-----------

Quando olhamos **a nível regional**, a contradição mostra-se correta (apenas a África Subsaariana contraria a tendência). A “inclinação positiva” global era uma ilusão ótica — paradoxo de Simpson — resultante da média de todas as regiões, mas regionalmente, a tendência do capital é clara. Ao olharmos para o cenário globalmente, acabamos por cair na narrativa do capital, que ganhos de produtividade equivale a ganhos salariais. A análise regional expõe a mentira — o *labour share* está basicamente estagnado em todos os lugares.



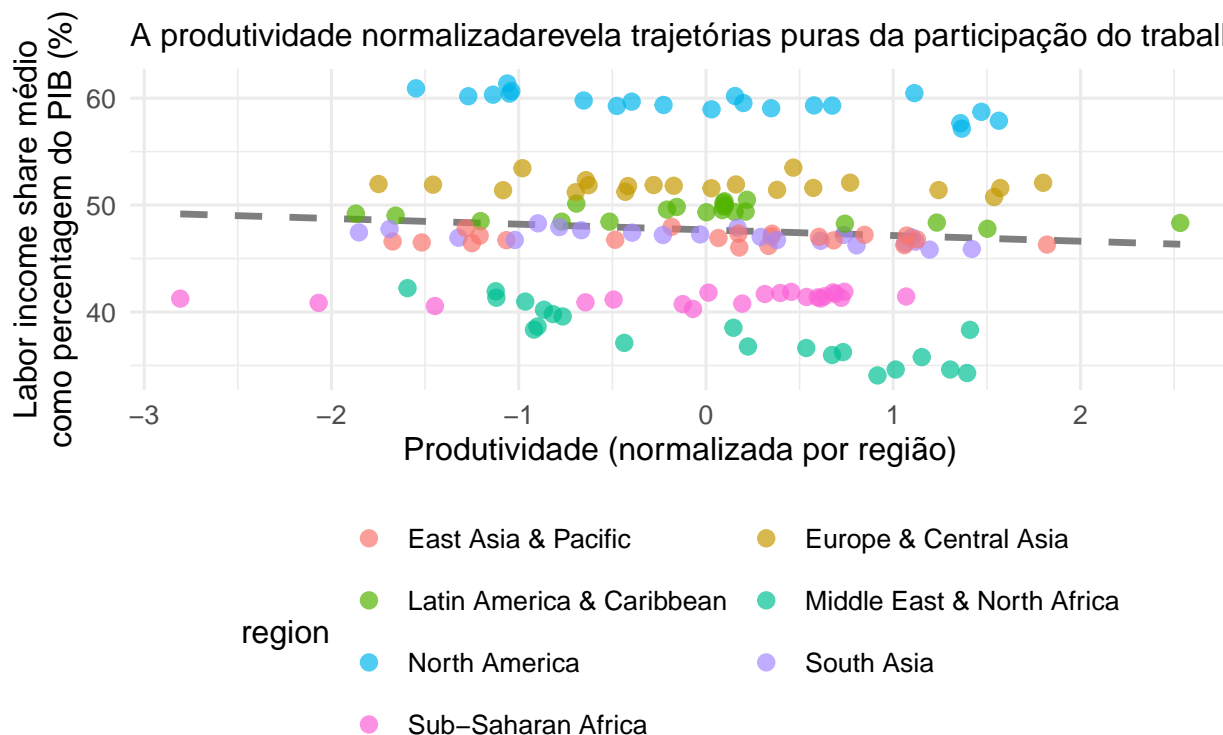
A linha global demonstra a ilusão de ótica resultante da mistura de regiões alta produtividade com com baixa produtividade. O declínio do *labour share* é uma característica estrutural do capitalismo em todo o

mundo.

Vamos normalizar a produtividade numa distribuição normal para uma última visualização.

Labour Share versus Produtividade normalizada

A produtividade normalizada revela trajetórias puras da participação do trabalho



Todas as regiões normalizadas com média 0, e desvio padrão 1. Foca-te na tendência do labour share.

2.3.2.4 Investimento direto estrangeiro versus labour share

Como os fluxos do IDE podem assumir valores negativos, aplicamos uma transformação logarítmica com sinal, preservando tanto a magnitude como a direção dos fluxos de capital, ao mesmo tempo que reduzimos a assimetria.

Sem o uso desta transformação, neste caso com a função logarítmica, estaríamos a juntar fluxos enormes com fluxos mais pequenos. Desta forma os valores maiores ficam comprimidos, diminuindo a variação, continuam a mostrar a sua dimensão, ao mesmo tempo que diminuímos os possíveis *outliers* que poderiam dificultar a análise do coeficiente de correlação Pearson.

A transformação logarítmica com sinal preserva a direção do fluxo, ao mesmo tempo que controla os valores monstruosos.

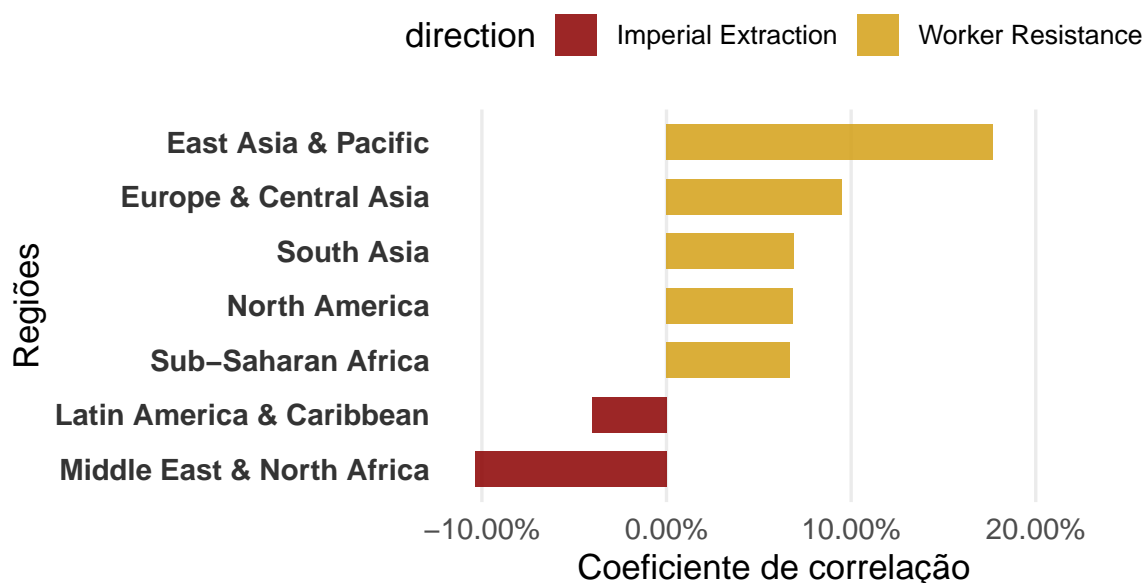
A correlação de Pearson usa a covariância que mede a direção do movimento conjunto deste caso da média dos fluxos do IDE com a média do *labour share*, ao longo do tempo, e divide pela multiplicação do desvio padrão dos dois. O desvio padrão mede o quão dispersos estão os dados em torno da sua média. Se estiverem muito dispersos, o valor é maior, e isso vai então diminuir o valor do coeficiente de correlação. No final, o coeficiente é um valor normalizado entre -1 e 1 . Se o valor for 0 , não existe correlação. Se for 1 , existe uma correlação perfeita onde um valor cresce juntamente com o outro, que é o caso oposto de se o

coeficiente for -1 , também uma correlação perfeita, mas um dos valores desce quando o outro sobe.

Para a esta correlação entre os fluxos IDE e o *labour share*, o coeficiente < 0 significa “mais extração, menos *labour share*” — sinal imperialista fundamental.

Esta visualização destaca correlações negativas no Sul Global (América Latina, Caraíbas, Médio Oriente e Norte de África) como evidência da destruição salarial através da extração imperialistas, enquanto correlações positivas em outros lugares expõem a colaboração entre burguesias.

Proxy de extração imperialista: correlações entre fluxos IDE e labour share por região



As correlações negativas na América Latina ($-4,1\%$) e na região Médio Oriente e Norte de África ($-10,4\%$) dão força à tese de que os fluxos IDE não representam “desenvolvimento”, mas sim a extração de valor dos trabalhadores para os centros imperialistas. O Leste Asiático e o Pacífico apresentam a correlação positiva mais forte entre fluxos IDE e *labour share* (Ásia Oriental $+17,7\%$), sugerindo resistência dos trabalhadores ou dinâmicas alternativas de capital como alianças capitalistas estatais. A correlação positiva da Ásia Oriental prova que a autogestão dos trabalhadores pode capturar/reverter os fluxos imperialistas.

Os valores do coeficiente de correlação, podem não parecer suficientemente fortes, até as regiões destacadas têm valores fracos. Mas esta é uma análise empírica. As fracas correlações não negam a extração imperialista. Correlações fracas são consistentes com um sistema no qual a extração de excedentes está estruturalmente incorporada.

2.4 Discussão

Na União Europeia, por exemplo, para conseguirem ter mais dinheiro para gastar em fins bélicos, vemos acontecer por vários países da federação, cortes orçamentais na segurança social e na educação e saúde pú-

blica, e com isso, reformas nas leis laborais com o intuito de aumentar a produtividade, que em troca, traz o aumento da precariedade do trabalhador e o aumento da desigualdade económica se a parcela salarial não aumentar com o aumento da produtividade. Por isso vemos na Europa várias greves gerais quase em simultâneo: 2 de dezembro na França, 28 de novembro e 12 de dezembro na Itália, de 24 a 26 de novembro na Bélgica, e aqui em Portugal a 11 de dezembro. Os sindicatos falham ao não relacionarem estes eventos, demonstrando que o problema não desaparece quando passamos as bordas do país, e que a luta e resistência é internacionalista, neste caso, é a nível europeu.

Na União Europeia, por exemplo, para conseguirem ter mais dinheiro para gastar em fins bélicos, vemos acontecer por vários países da federação, cortes orçamentais na segurança social e na educação e saúde pública, e com isso, reformas nas leis laborais com o intuito de aumentar a produtividade, que em troca, traz o aumento da precariedade do trabalhador e o aumento da desigualdade económica se a parcela salarial não aumentar com o aumento da produtividade. Por isso vemos na Europa várias greves gerais quase em simultâneo.

3 Conclusão

Referências Bibliográficas

- Apache Software Foundation. (s.d.). *Apache Airflow*. <https://airflow.apache.org/>
- Bescond, D., Villarreal-Fuentes, M., & of Statistics, I. D. (2025). *Rilostat: R Interface to ILOSTAT Open Data* (Versão 2.3.4). <https://cran.r-project.org/web/packages/Rilostat/index.html>
- DuckDB. (2025). <https://duckdb.org/>
- Fireship. (2025). <https://www.youtube.com/c/Fireship>
- Nikšić, H. (s.d.). *GNU Wget*. <https://www.gnu.org/software/wget/>
- OpenRefine Community. (2025). *OpenRefine*. <https://openrefine.org/>
- Perplexity AI. (s.d.). Perplexity AI, Inc. <https://www.perplexity.ai/>
- Prefect Technologies. (s.d.). *Prefect*. <https://www.prefect.io/>
- R Core Team. (2025). *An Introduction to R*. Versão 4.5.2. R Foundation for Statistical Computing, Vienna, Austria. <https://cran.r-project.org/doc/manuals/r-release/R-intro.html>
- Spotify AB. (s.d.). *Luigi: Workflow Scheduling System*. https://luigi.readthedocs.io/en/stable/running_luigi.html
- TimescaleDB. (2025). <https://github.com/timescale/timescaledb>
- Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., & Posit Software, P. (2023). *dplyr: A Grammar of Data Manipulation* (Versão 1.1.4). <https://cran.r-project.org/web/packages/dplyr/index.html>
- Wickham, H., & Posit Software, P. (2025). *stringr: Simple, Consistent Wrappers for Common String Operations* [R package version 1.6.0]. <https://cran.r-project.org/web/packages/stringr/index.html>
- Wickham, H., Vaughan, D., Girlich, M., Ushey, K., & Posit Software, P. (2025). *tidyr*. <https://tidyverse.org/>

Licença

Este documento está licenciado sob uma Licença Creative Commons Atribuição–Partilha nos Mesmos Termos 4.0 Internacional (CC BY-SA 4.0).

O código fonte (ficheiros `.tex`, `.bib`, `Makefile`, etc.) utilizado para produzir este relatório está licenciado sob a GNU Affero General Public License v3.0 (AGPL v3).

A Indicadores

A.1 PIB

GDP

GDP (current US\$)

Gross domestic product is the total income earned through the production of goods and services in an economic territory during an accounting period. It can be measured in three different ways: using either the expenditure approach, the income approach, or the production approach. This indicator is expressed in current prices, meaning no adjustment has been made to account for price changes over time. This indicator is expressed in United States dollars.

A.2 Participação do rendimento do trabalho

Labour (income) share

A.3 Índice de Gini

Gini index

Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.

A.4 Investimento direto estrangeiro

Investimento direto estrangeiro, entradas líquidas (Balança de Pagamentos, US\$ atuais)

Foreign direct investment, net inflows (BoP, current US\$)

Foreign direct investment refers to direct investment equity flows in the reporting economy. It is the sum of equity capital, reinvestment of earnings, and other capital. Direct investment is a category of cross-border investment associated with a resident in one economy having control or a significant degree of influence on the management of an enterprise that is resident in another economy. Ownership of 10 percent or more of the ordinary shares of voting stock is the criterion for determining the existence of a direct investment relationship. Data are in current U.S. dollars.

A.5 Profit sharing

A.6 Real wages

B Justificação das escolhas tecnológicas

Eu uso Linux e não Microsoft Windows já faz 5 anos. Por várias razões, mas tudo tem a mesma raiz do problema: Windows é *software* proprietário.

B.1 *Software* livre

Eu vou, sempre que conseguir, usar *free, libre, and open source software* (FLOSS). Muitas vezes tenho que ceder aos monopólios do *software*, porque não o fazer prejudica-me na vida fora do digital. Por exemplo, parece que muita da infraestrutura informática do politécnico depende da Microsoft, a começar pelos correios eletrónicos disponibilizados para os alunos. Eu sendo aluno, tenho que o usar porque senão perco acesso à informação importante para completar a minha formação. E é assim que estes oligarcas digitais criam dependências nos seus *software*. Mas eles são capazes de se introduzir em muito mais do que apenas na vida académica e nos estudos. Quer dizer, caraças, eles metem-se em tudo o que conseguem, desde que haja lucro sendo feito.

Na pedagogia e ensino, devia ser ilegal o uso de *software* proprietário, já que este vai contra todos os fundamentos da educação. *Software* livre tem que ser a norma. O *software* que deixa o estudante aprender e estudar o *software* que usa, executar o *software* da forma que lhe for mais conveniente, modificar para as suas necessidades, e conseguir partilhar a sabedoria que adquiriu com qualquer um.

Acho que já deu para ter uma ideia do porquê da minha indignação. E para minha surpresa, esta cadeira pede para o aluno usar Microsoft Excel, Microsoft Power BI e Microsoft SQL Server. Primeiro, que eu nem consigo usar alguns desses *softwares* no meu portátil sem ter que ter o Windows como sistema operativo — isso é que nem pensar. E depois que existem alternativas, e boas (senão melhores) que são livres. Vou mencionar OpenRefine e Apache Superset que fazem tudo o que era necessário. Não os usei porque, igual às ferramentas da Microsoft, esses não oferecem uma forma simples de reproduzir as mesmas ações em qualquer outra máquina.

B.2 Dados Abertos

Falando mais sobre a cadeira, uns tópicos que eu acho que deviam ser pelo menos mencionados é o dos dados abertos mencionando o esquema da implementação das 5 estrelas e os princípios *Findable, Accessible,*

Interoperable, Reusable (FAIR). Mas claramente não querem saber disso. Incentivam a usar o formato proprietário do Excel (duas estrelas). Não querem nos ensinar a explorar, aprender, adquirir um pensamento crítico independente, mas sim a consumir a droga do *software* proprietário, e aceitar que tem que ser assim, que estas são as ferramentas corretas.

Não querem ensinar como aprender, mas sim o que aprender. A resistência em adotar a educação aberta decorre do desejo de controlar o que aprendemos. E nem tentam esconder mais isto. É triste perguntar a razão de estarmos a aprender algo e a resposta ser “porque é o que as empresas querem”.