

Apresentado por:



Google Cloud

Accelerate

Relatório

State of

DevOps

2023



Patrocinadores Premiere



Deloitte.

Qarik

Sumário

Início

Resumo executivo	03
Conceitos e medidas	06

Capítulo 1

Sua equipe está à altura?	10
---------------------------------	----

Capítulo 2

Foco nos usuários	17
-------------------------	----

Capítulo 3

Capacidades técnicas que predizem o desempenho	20
---	----

Capítulo 4

O papel fundamental da documentação	27
---	----

Capítulo 5

A confiabilidade aumenta o desempenho	31
--	----

Todas as citações foram coletadas em 27 de setembro de 2023

Capítulo 6

A infraestrutura flexível é a chave do sucesso	38
---	----

Capítulo 7

Nada disso funciona sem investir na cultura	45
--	----

Capítulo 8

A importância de "como", "quando", "por quê" e "quem"	51
--	----

Posfácio

Considerações finais	57
Agradecimentos	58
Autores	59
Metodologia	62
Informações demográficas e firmográficas	72
Os modelos	81
Sugestões de leitura	91
Apêndice	92

Resumo executivo

Há quase uma década, o programa de pesquisa DORA investiga capacidades e medidas de organizações de alto desempenho baseadas em tecnologia. Já ouvimos mais de 36.000 profissionais de organizações de diversos tamanhos e setores. Agradecemos por compartilhar seus insights!

A DORA busca entender a relação entre formas de trabalho (ou seja, capacidades) e resultados: conquistas importantes que são relevantes para a organização e as pessoas que atuam nela. Esta pesquisa segue uma avaliação estatística rigorosa e não se limita a nenhuma plataforma específica (consulte [Metodologia](#)).

Esperamos que estes insights mostrem a líderes e profissionais onde é possível ter mais impacto.

A pesquisa deste ano investigou três resultados importantes e as capacidades correspondentes:

- **Performance organizacional:** a organização não deve gerar somente receita, mas também valor para clientes e para a comunidade como um todo.
- **Desempenho da equipe:** a habilidade de uma equipe de serviço ou aplicativo gerar valor, inovar e colaborar.
- **Bem-estar dos funcionários:** as estratégias que uma organização ou equipe adotam devem beneficiar os funcionários, reduzindo o esgotamento, aumentando a satisfação no trabalho e melhorando a capacidade das pessoas de alcançar resultados valiosos (ou seja, ter produtividade).

A pesquisa também investigou formas ou medidas de desempenho que costumamos definir como objetivos finais:

- **Desempenho de entrega de software:** as equipes podem mudar sistemas tecnológicos com segurança, rapidez e eficiência.
- **Desempenho operacional:** o serviço oferece uma experiência confiável aos usuários.



Principais descobertas

» Desenvolva uma cultura saudável

A cultura é a base para aprimorar capacidades técnicas, impulsionar o desempenho técnico, alcançar metas organizacionais de performance e ajudar os funcionários a serem bem-sucedidos. Equipes com culturas generativas têm uma performance organizacional 30% mais alta.

» Crie pensando nos usuários

O foco no usuário pode guiar e impulsionar o aprimoramento de todas as capacidades técnicas, processuais e culturais investigadas na nossa pesquisa. Independentemente da agilidade e do sucesso da implantação, as equipes precisam pensar no usuário para que o trabalho tenha efeito. Equipes que focam no usuário têm uma performance organizacional 40% mais alta.

» Aumente o desempenho de entrega de software com revisões de código mais ágeis

Acelerar as revisões de código é uma das formas mais eficientes de melhorar o desempenho de entrega de software. Equipes que fazem revisões de código mais rápidas têm um desempenho de entrega de software 50% mais alto.

» Tenha uma documentação de qualidade para intensificar as capacidades técnicas

Ter uma documentação de alta qualidade intensifica o impacto das capacidades técnicas na performance organizacional. Por exemplo, estima-se que o desenvolvimento baseado em linha principal tenha um impacto 12,8 vezes maior na performance organizacional quando há uma documentação de alta qualidade (em comparação com uma de baixa qualidade).

» Use a nuvem para ter uma infraestrutura mais flexível

Uma das vantagens da computação em nuvem é a flexibilidade que ela traz para a infraestrutura. Por exemplo, usar uma nuvem pública torna a infraestrutura 22% mais flexível em comparação a infraestruturas sem nuvem. Essa flexibilidade gera um aumento de 30% da performance organizacional em comparação a infraestruturas inflexíveis. Para aproveitar ao máximo a nuvem, é fundamental usufruir dos diferenciais e das capacidades que ela oferece, especificamente a flexibilidade que traz para a infraestrutura.

» Equilibre velocidade de entrega, desempenho operacional e foco no usuário

Para ter a melhor performance organizacional possível, você precisa de uma entrega de software sólida e um desempenho operacional excelente. O equilíbrio entre esses dois elementos, aliado ao foco no cliente, é ideal para ter resultados organizacionais melhores e ainda aumentar o bem-estar dos funcionários.

» Distribua o trabalho de maneira justa

Pessoas que se consideram sub-representadas e mulheres ou pessoas de gênero autodeterminado têm níveis mais altos de esgotamento. Esse efeito pode ter várias causas relacionadas ao sistema e ao ambiente. Por isso, não é surpresa que os participantes da pesquisa que assumem trabalhos mais repetitivos são mais propensos a ter níveis altos de esgotamento, e membros de grupos sub-representados têm mais chance de assumir trabalhos repetitivos. Participantes sub-representados têm 24% mais esgotamento do que os não sub-representados. Participantes sub-representados fazem 29% mais trabalhos repetitivos do que os não sub-representados. Mulheres e pessoas de gênero autodeterminado fazem 40% mais trabalhos repetitivos do que homens.

Como aplicar os insights da DORA ao seu contexto

Equipes que adotam uma mentalidade e uma prática contínua voltadas a melhorar tendem a se beneficiar mais.¹ A DORA pode ajudar a influenciar suas próprias iniciativas de aprimoramento.

Para aproveitar tudo que esta pesquisa tem a oferecer, considere o contexto da sua equipe e dos seus usuários. Por exemplo, mencionamos antes que equipes com revisões de código mais rápidas têm um desempenho de entrega de software 50% mais alto. Mas seu desempenho de entrega de software provavelmente não vai melhorar se as revisões de código já forem rápidas e o gargalo da velocidade estiver em outra parte do sistema. Para aplicar a pesquisa ao seu contexto, os profissionais precisam conversar sobre como o trabalho é feito hoje. Essas conversas podem melhorar a empatia, a colaboração e a compreensão das motivações de cada participante.

O trabalho de aprimoramento nunca acaba. Encontre um gargalo no seu sistema, resolva e repita o processo. As comparações mais importantes surgem analisando o mesmo aplicativo várias vezes ao longo do tempo, não quando você olha outros aplicativos, organizações ou setores.

Métricas e medidas

Métricas e painéis ajudam as equipes a monitorar o progresso e corrigir o direcionamento.

Profissionais e líderes querem melhorar a performance organizacional, o desempenho da equipe e o bem-estar. Mas medir isso (ou a entrega de software) não é o objetivo.

A fixação em métricas de desempenho pode causar comportamentos ineficazes. Uma forma melhor de alcançar o sucesso é investindo em capacidades e aprendizado. Equipes que aprendem mais se aprimoram mais.

Para aprimorar, é importante não se isolar

Podemos aprender com a experiência de cada pessoa, e um espaço excelente para compartilhar e aprender sobre iniciativas de aprimoramento é o site da Comunidade DORA <https://dora.community> (em inglês).

¹2022 Accelerate State of DevOps Report. <https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report.pdf#page=7>



Conceitos e medidas

Nesta seção, vamos descrever os conceitos que a DORA tenta medir.¹ Esses elementos são a base do relatório e dos nossos modelos. Estas seções mostram os ingredientes que usamos para criar o relatório. Por isso, é importante que nós, os autores, deixemos claro o que são esses conceitos e falemos sobre eles de forma consistente. As tabelas a seguir servem para esclarecer e definir uma terminologia compartilhada entre leitores e autores.

Como muitos conceitos deste relatório são multifacetados, às vezes usamos vários indicadores para capturá-los. Uma maneira de avaliar se esses conceitos foram capturados com sucesso é fazendo análises fatoriais exploratórias e confirmatórias. Leia mais sobre esses processos na seção [Metodologia](#). Depois de avaliar a metodologia de medição, distribuímos as pontuações em uma escala de 0 a 10, em que 0 representa a ausência total e 10 significa que um conceito está sempre presente. Acreditamos que isso ajuda a padronizar como falamos sobre esses conceitos e a comparar dados ao longo dos anos.

Cada conceito discutido será acompanhado das seguintes informações:

- Um ícone que ajuda a comunicar um significado e mostra quando um capítulo deve ser usado como referência.
- A pontuação média do conceito na amostra (a média).
- Os limites do intervalo interquartil (IIQ). Esses limites são os dois números ($o\ 25^{\circ}$ e $o\ 75^{\circ}$ percentil) entre os quais estão 50% dos dados, assim ficam mais fácil mostrar a distribuição das respostas.
- O valor médio de um conjunto de dados (a mediana). Se ela for muito diferente da média, os dados podem estar distorcidos.
- Uma descrição do conceito e como ele foi medido.

¹As perguntas de pesquisa usadas na nossa análise estão disponíveis em <https://dora.dev>

Resultados principais

Os resultados principais são os objetivos que nós acreditamos que pessoas, equipes ou organizações estão tentando alcançar (como performance organizacional) ou evitar (como esgotamento). Acreditamos que medidas são importantes para que as pessoas possam avaliar a si mesmas, suas equipes e suas organizações.

 Performance organizacional	 Desempenho da equipe	 Desempenho de entrega de software
Média 6,3	IIQ 5-8	Mediana 6,3
Organizações com alta performance têm mais clientes, lucros mais altos e participações maiores no mercado para seu produto ou serviço principal.	Em equipes com alto desempenho, os membros se adaptam a mudanças, confiam uns nos outros, trabalham de modo eficiente, inovam e colaboram.	Estas quatro métricas definem a velocidade e a estabilidade da entrega de software:
 Desempenho operacional	 Metas de confiabilidade	<ul style="list-style-type: none">• Frequência de implantação• Tempo de lead das mudanças• Taxa de falha nas mudanças• Tempo de recuperação após falha ao implantar
Média 6,2	IIQ 5-7,5	Mediana 6,3
Até que ponto um serviço é capaz de atender às expectativas dos usuários, incluindo aspectos como disponibilidade e desempenho.	Até que ponto um serviço cumpre as metas a que se propõe com base em aspectos como disponibilidade, desempenho e correção.	

Bem-estar é a soma do nível de esgotamento, produtividade e satisfação no trabalho

 Esgotamento	 Produtividade	 Satisfação no trabalho
Média 4,1	IIQ 2-6	Mediana 4
O peso psicológico e físico, além de quanto valor e significado cada pessoa atribui ao próprio trabalho. O esgotamento gera cinismo. ²	Uma pessoa produtiva realiza tarefas condizentes com suas habilidades, gera valor e trabalha de modo eficiente.	Uma pergunta única que pede para o participante considerar tudo e classificar como se sente em relação ao trabalho no geral. ³

² Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry. 2016 Jun;15(2):103-11. doi: 10.1002/wps.20311. PMID: 27265691; PMCID: PMC4911781.

³ Warr, P., Cook, J., & Wall, T. "Scales for the measurement of some work attitudes and aspects of psychological well-being". Journal of Occupational Psychology, 52(2), 1979. 129-148. <https://doi.org/10.1111/j.2044-8325.1979.tb00448.x>

Processos e capacidades técnicas

Atividades, práticas ou estados que podem surgir em uma equipe ou organização.

Em outras palavras, o que as equipes fazem ou são.

 Contribuição da inteligência artificial Média IIQ Mediana 3,3 0,3-6,3 2,4 Quanto a inteligência artificial contribui para uma variedade de tarefas técnicas.	 Arquitetura com acoplamento flexível Média IIQ Mediana 6,4 4,7-8,3 6,7 Software que pode ser gravado, testado e implantado de forma independente.	 Integração continua Média IIQ Mediana 6,9 5-8,9 7,8 A prática de automaticamente criar e testar mudanças no software.
 Entrega contínua Média IIQ Mediana 7,0 5,7-8,7 7,3 A capacidade de colocar vários tipos de mudanças (como novos recursos, alterações nas configurações, correções de bugs e experimentos) em produção ou nas mãos dos usuários de forma segura, rápida e sustentável. ⁴	 Velocidade de revisão de código Média IIQ Mediana 6,5 6-8 6 Item único que avalia o tempo entre a solicitação de envio até a revisão de alteração do código.	 Documentação Média IIQ Mediana 5,8 3,8-7,9 6,25 A qualidade do conteúdo gravado que as pessoas da organização criam e usam no trabalho diário.
 Práticas de confiabilidade Média IIQ Mediana 5,9 3,9-8,3 6,1 Atividades e práticas que as equipes usam para melhorar o desempenho operacional dos serviços.	 Desenvolvimento baseado em linha principal Média IIQ Mediana 5,6 3,9-7,8 5,6 A prática de fazer mudanças pequenas e frequentes que costumam ser incorporadas à ramificação do código principal no sistema de controle de versões.	 Infraestrutura flexível Média IIQ Mediana 6,6 5-8,3 7,3 Infraestrutura dimensionável que é elástica, acessível e mensurável. ⁵

⁴“What is Continuous Delivery” <https://continuousdelivery.com/>

⁵ National Institute of Standards and Technology (2018) NIST The NIST Definition of Cloud Computing. Disponível em <https://csrc.nist.gov/pubs/sp/800-145/final>

Aspectos culturais

Não é fácil definir o que é cultura, mas podemos dizer que são as normas predominantes (como a flexibilidade), a orientação prevalente (como a centralidade do usuário) e o clima (como a estabilidade organizacional) do espaço de trabalho.

 Distribuição do trabalho	 Flexibilidade	 Estabilidade do emprego
Média 5,8	IIQ 3,8-7,9	Mediana 5,8
Processos formais para ajudar os funcionários a terem uma distribuição igualitária de tarefas na equipe.	Como, onde e quando uma pessoa trabalha nas tarefas. ⁶	Uma medida de um item só que pergunta com que frequência as pessoas se preocupam com a estabilidade do emprego. Pontuações maiores indicam menos preocupação.
 Estabilidade organizacional	 Compartilhamento de conhecimentos	 Centralidade do usuário
Média 7,2	IIQ 6,7-8,3	Mediana 8,3
Uma medida de um item só que pergunta até que ponto um ambiente de trabalho é estável ou instável para os funcionários.	Como ideias e informações se espalham por uma organização. Os membros da equipe respondem às perguntas uma vez e disponibilizam as informações para os outros. As pessoas não precisam esperar pelas respostas. ⁷	Entender e incorporar as necessidades e objetivos dos usuários para melhorar produtos e serviços. ⁸
 Cultura organizacional de Westrum		
Média 7,3	IIQ 6,1-8,6	Mediana 7,8
Como uma organização tende a reagir a problemas e oportunidades. Há três tipos de culturas: generativa, burocrática e patológica. ⁹		

⁶Shifrin, Nicole V. e Jesse S. Michel. "Flexible work arrangements and employee health: A meta-analytic review". Work & Stress 36, nº. 1, 2022. 60-85

⁷"2022 Developer Survey" <https://survey.stackoverflow.co/2022#overview>

⁸Kersten, Mik. Project to Product: How to survive and thrive in the age of digital disruption with the flow framework (IT Revolution, 2018), 54. <https://itrevolution.com/product/project-to-product/>

⁹Westrum R. "A typology of organisational cultures." BMJ Quality & Safety, 2004. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1765804/>

Sua equipe está à altura?

A lei de Goodhart: quando uma medida se torna uma meta, ela deixa de ser uma boa medida.¹

Pontos principais

A primeira etapa para melhorar a performance é definir o valor de referência atual para o desempenho de entrega de software, o desempenho operacional e a centralidade do usuário de um aplicativo. Essas medidas ajudam as equipes a avaliar como estão e fornecem um bom indicativo da mudança ao longo do tempo.

Mas elas não servem para aprimorar a equipe. Depois de definir o valor de referência, é importante avaliar os pontos fortes da empresa, incluindo pessoas, processos e capacidades técnicas, para identificar por que o progresso está estagnado.² Depois, as equipes precisam de tempo e espaço para alinhar, testar e reavaliar. Repetir esse processo ajuda as equipes a adotar uma mentalidade e práticas de aprimoramento contínuo.

Cuidado com estas e outras armadilhas quando usar estas comparações:

- **Fazer comparações que não correspondem.** Não compare aplicativos com base apenas nestes clusters. Comparações assim não consideram o contexto de cada aplicativo e podem prejudicar a busca pelo aprimoramento.
- **Definir métricas como metas.** Ignorar a lei de Goodhart e fazer afirmações vagas com “todos os aplicativos precisam ter uma performance de ‘elite’ até o final do ano” aumenta a probabilidade das equipes manipularem as métricas.
- **Eleger uma métrica como prioridade.** Tentar medir sistemas complexos com “uma métrica que importa”. Usar uma combinação de métricas para aprofundar o entendimento.³
- **Usar métricas com um escopo muito específico.** As pessoas tendem a medir o que é mais mensurável, não aquilo que é mais significativo.
- **Usar o setor como desculpa para não se aprimorar.** Por exemplo, algumas equipes de setores altamente regulamentados usam as normas como motivo para não desafiar o status quo.

Para saber mais sobre nossas descobertas e recomendações para ter uma mentalidade e uma prática de aprimoramento contínuo, confira “How to transform” em: dora.dev/devops-capabilities/cultural/devops-culture-transform/ (em inglês)

¹Strathern, Marilyn (1997). “‘Improving ratings’: audit in the British University system”. European Review. John Wiley & Sons. 5 (3): 305–321. doi:10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4. S2CID 145644958.

²Este relatório e os recursos listados no “Capacity catalog” (<https://dora.dev/devops-capabilities/>, em inglês) podem ser úteis.

³Forsgren, N., Storey, M-A., et al. “The SPACE of Developer Productivity: There’s more to it than you think.”, 2021. <https://queue.acm.org/detail.cfm?id=3454124>

Introdução

Todo ano realizamos uma ou mais análises de clusters para encontrar tendências comuns entre os aplicativos. Recomendamos que você use essas análises para entender como você está em relação aos demais, mas não se prenda a essas comparações. As melhores comparações são as feitas com o mesmo aplicativo ao longo do tempo em vez de entre aplicativos diferentes, que sempre terão contextos distintos.

As equipes criam software para os usuários, que são os que julgam a confiabilidade e utilidade do serviço. Equipes que focam nas necessidades dos usuários são as mais bem equipadas para desenvolver uma solução ideal. E as equipes que aliam o foco no usuário ao desempenho de entrega de software e das operações são as mais preparadas para fazer isso do jeito certo.

Equipes que focam nas necessidades dos usuários criam a solução ideal E do jeito certo.

Resultados

Desempenho de entrega de software

Usamos as seguintes medidas para avaliar o desempenho de entrega de software:

- **Tempo de lead das mudanças:** quanto tempo leva para um compromisso de mudança virar uma implantação.
- **Frequência de implantação:** frequência de envio das mudanças para produção.
- **Taxa de falha em mudanças:** com que frequência uma implantação introduz uma falha que exige intervenção imediata.
- **Tempo de recuperação após falha ao implantar:** quanto tempo é necessário para se recuperar de uma falha de implantação.

Uma abordagem comum para melhorar as quatro medidas é reduzir o tamanho do lote de mudanças de um aplicativo.⁴ Mudanças menores são mais fáceis de lidar e avançar ao longo do processo de entrega. Elas também facilitam o processo de recuperação em caso de falha. As equipes devem fazer as menores mudanças possíveis para que o processo de entrega seja rápido e estável. Isso melhora a velocidade e a estabilidade das mudanças.

⁴Muitas vezes, um recurso pode ser dividido em várias alterações entregues de forma independente. Nossas medidas de desempenho de entrega de software avaliam as mudanças feitas em um aplicativo ou serviço.

Este ano, refinamos as medidas de desempenho de entrega de software. Leia mais sobre essas mudanças em “Refinamento das medidas de desempenho de entrega de software” no [Apêndice](#).

Aqui está um panorama dos participantes da pesquisa deste ano em relação ao desempenho de entrega de software:

Nível de desempenho	Frequência de implantação	Tempo de lead das mudanças	Taxa de falha nas mudanças	Tempo de recuperação após falha ao implantar	% de participantes
Elite	Sob demanda	Menos de um dia	5%	Menos de uma hora	18%
Alto	Entre uma vez por dia e uma vez por semana	Entre um dia e uma semana	10%	Menos de um dia	31%
Médio	Entre uma vez por semana e uma vez por mês	Entre uma semana e um mês	15%	Entre um dia e uma semana	33%
Baixo	Entre uma vez por semana e uma vez por mês	Entre uma semana e um mês	64%	Entre um mês e seis meses	17%



Desempenho operacional

Avaliamos o desempenho operacional perguntando com que frequência o serviço dos participantes:

- Recebe relatórios de usuários finais insatisfeitos com a confiabilidade do sistema.
- Não está disponível ou tem um desempenho incorreto ou mais lento do que o previsto.

Para saber como o desempenho operacional prediz a performance organizacional, consulte o [Capítulo 5: A confiabilidade aumenta o desempenho](#).

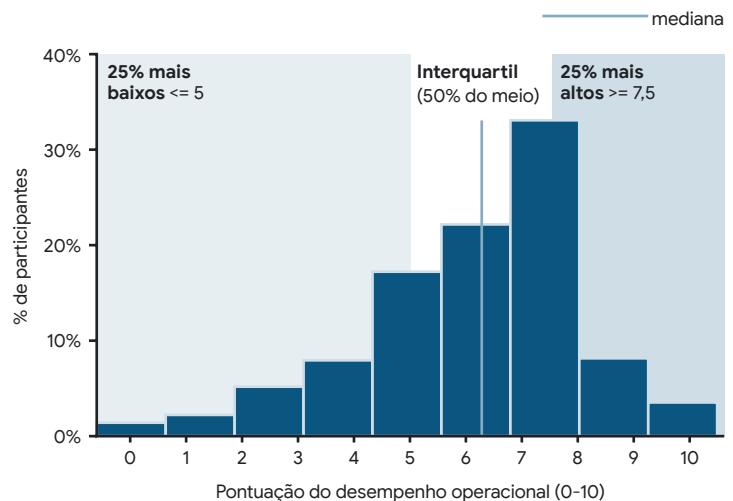
Centralidade do usuário

Um aplicativo ou serviço centrado no usuário é criado pensando no usuário final. Esse modelo de desenvolvimento de produto exige entender o que os usuários precisam e incorporar isso à estratégia do produto.

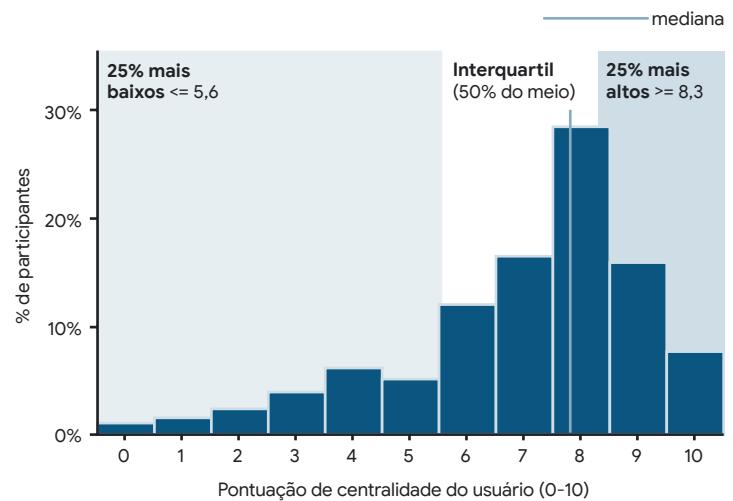
Para avaliar a centralidade do usuário, perguntamos aos participantes até que ponto eles concordam com as seguintes afirmações:

- A equipe entende bem o que os usuários querem fazer.
- O sucesso da equipe é avaliado com base no valor que ela agrega à organização e aos usuários do aplicativo.
- As especificações (por exemplo, planejamento de requisitos) são continuamente revistos e priorizados de novo conforme indicadores dos usuários.

Aqui está um panorama dos participantes da pesquisa deste ano em relação ao desempenho operacional:



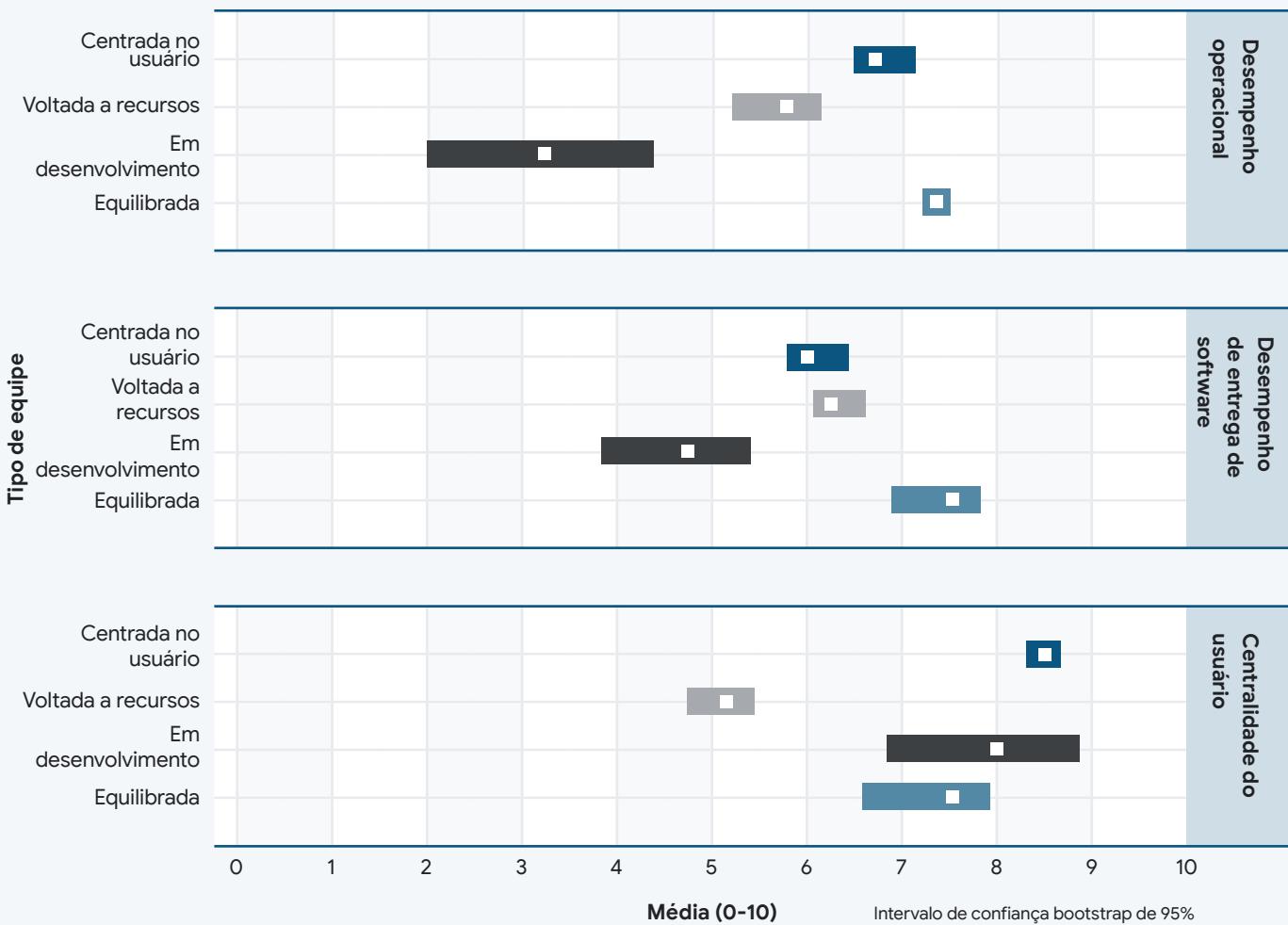
Aqui está um panorama dos participantes da pesquisa deste ano em relação à centralidade do usuário:



Cluster de tipos de equipes

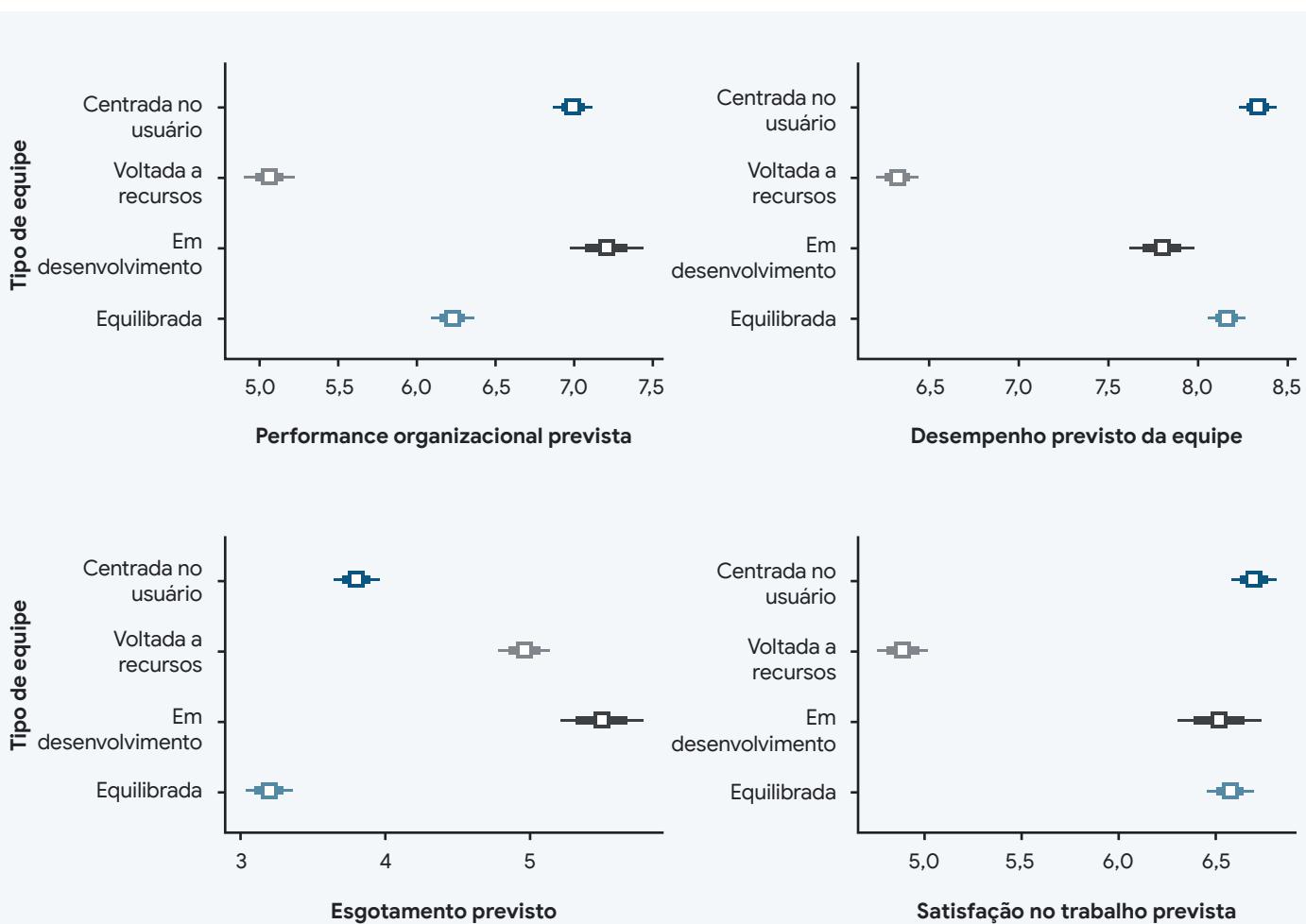
Com base na comparação entre desempenho de entrega de software, desempenho operacional e centralidade do usuário é possível definir quatro tipos de equipes. Como todas as medidas usadas para criar esses tipos de equipes, eles estão no âmbito de aplicativos ou serviços.

Os tipos de equipes são Centrada no usuário, Voltada a recursos, Em desenvolvimento e Equilibrada



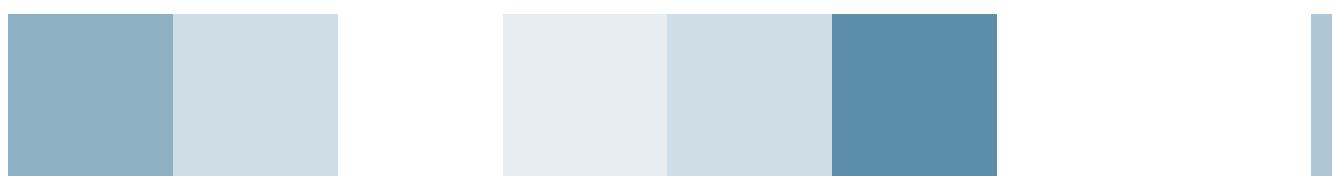
As métricas de desempenho que temos discutido são como comandos que uma organização ou equipe pode ajustar para mudar a performance organizacional, o desempenho da equipe e o bem-estar das pessoas na equipe.

Os gráficos abaixo mostram os resultados de desempenho previstos para cada tipo de equipe.



*O ponto representa a média estimada para o tipo de equipe. Intervalo amplo que engloba 66% das simulações. Intervalo estreito que engloba 89% das simulações.

Cada tipo de equipe tem características específicas, compõe uma proporção significativa dos participantes e tem resultados diferentes. É provável que sua própria equipe não se encaixe perfeitamente em apenas um desses tipos, e não esperamos que o tipo da equipe continue sendo o mesmo ao longo do tempo.



O que esses resultados significam?

» Equipe centrada no usuário

Esse tipo de equipe foca principalmente nas necessidades do usuário. Aliado a um bom desempenho operacional e de entrega de software, esse foco é um preditor de que a performance organizacional está altíssima. No entanto, equipes assim têm um pouco mais de esgotamento do que as equilibradas. Aumentar o desempenho de entrega de software e/ou das operações pode ser a melhor forma de reduzir o esgotamento nessas equipes.

» Equipe voltada a recursos

Esse tipo de equipe prioriza a entrega de recursos. Um foco implacável na entrega pode fazer a equipe não prestar atenção às necessidades dos usuários, conforme demonstrado pelos baixos números de centralidade do usuário e desempenho operacional. Equipes desse tipo têm os mais altos níveis de esgotamento e os mais baixos de satisfação no trabalho, desempenho da equipe e performance organizacional. Os funcionários preferem entregar valor, não só recursos. A recomendação para equipes voltadas a recursos é refletir sobre as necessidades dos usuários para agregar mais valor aos recursos enviados.

» Equipe em desenvolvimento

Equipes desse tipo têm uma boa performance organizacional porque focam nas necessidades dos usuários do aplicativo. No entanto, pode ser que elas ainda estejam em busca do espaço ideal no mercado para as capacidades técnicas ou o produto delas. Essas equipes costumam fazer parte de organizações menores. Elas têm desempenho operacional e desempenho de entrega de software mais baixos, e as equipes que trabalham nesses aplicativos têm mais esgotamento do que as equilibradas ou centradas no usuário. Essas equipes podem ter processos de grande porte ou tarefas trabalhosas que podem ser automatizadas para melhorar o desempenho de entrega de software e de operações.

» Equipe balanceada

Equipes desse tipo têm uma abordagem equilibrada e sustentável. Elas usam tecnologia de forma sustentável para ter uma boa performance organizacional e boa satisfação no trabalho. Essas equipes também relatam os níveis mais baixos de esgotamento. Essas equipes ajustaram as próprias capacidades para ter um bom desempenho nas três medidas. Aumentar a centralidade do usuário pode ser a melhor forma de aprimorar a performance organizacional.

O foco nos usuários prediz a performance organizacional

Pontos principais

As organizações podem se atualizar sobre as últimas tendências de tecnologia e gestão conforme tentam melhorar a produtividade dos desenvolvedores e a performance organizacional. Nossa pesquisa mostra que uma abordagem de criação de aplicativos e serviços centrada no usuário é um dos maiores preditores da performances organizacional geral. Para melhorar a performance, entenda melhor seus usuários, faça iterações para ajustar e incorpore o feedback deles.

Introdução

O movimento de DevOps começou como uma forma de incentivar a colaboração entre equipes de desenvolvimento e operações para oferecer mais valor ao usuário. Esse alinhamento teve sucesso logo no início e viabilizou ideias e capacidades de DevOps que vão além desses dois departamentos. Hoje, as organizações de alto desempenho baseadas em tecnologia reconhecem a importância do alinhamento de equipes para alcançar metas organizacionais.

Investigamos três características críticas do foco no usuário:

- Quanto as equipes entendem as necessidades dos usuários.
- Quanto a equipe está alinhada para atender às necessidades dos usuários.
- Como o feedback dos usuários é usado no momento de definir as prioridades do trabalho.

Equipes com um forte foco no usuário têm
40%
mais performance organizacional

Resultados

Descobrimos que o desenvolvimento de software centrado no usuário causa um aumento significativo na performance. As organizações podem ter uma série de benefícios quando colocam o usuário em primeiro lugar. O feedback dos usuários ajuda as equipes a priorizar projetos e criar produtos e serviços que atendam às necessidades deles. Essa abordagem aprimora a experiência dos usuários, melhora a satisfação deles e aumenta a receita.

Efeito da centralidade do usuário em...

Performance organizacional		Aumento substancial
Desempenho da equipe		Aumento substancial
Desempenho de entrega de software		Aumento pequeno
Desempenho operacional		Aumento substancial
Desenvolvimento baseado em linha principal		Aumento substancial
Práticas de confiabilidade		Aumento substancial
Integração contínua		Aumento substancial
Entrega contínua		Aumento substancial
Arquitetura com acoplamento flexível		Aumento substancial
Esgotamento		Diminuição pequena*
Satisfação no trabalho		Aumento substancial
Produtividade		Aumento substancial

*Diminuir o esgotamento é algo positivo!

O que esses resultados significam?

Focar nas necessidades dos usuário é um ótimo preditor da performance organizacional geral. Para ter um bom foco nos usuários, são necessários incentivos, alinhamento e modelos de trabalho adequados. O foco nos usuários tem o potencial de guiar e impulsionar todas as capacidades técnicas, processuais e culturais que investigamos na pesquisa.

Aqui está uma análise de como esses resultados podem afetar várias equipes da sua organização:

Equipes de entrega e desenvolvimento de produtos

» O foco no usuário ajuda a garantir que as equipes de entrega e desenvolvimento de produtos criem o que os usuários querem, de preferência de forma sustentável. Equipes equilibradas fazem exatamente isso. Elas têm um bom nível de entrega, operações e performance organizacional, com um foco sólido nas necessidades dos usuários. Os membros dessas equipes entendem claramente as necessidades dos usuários e conseguem ajustar os planos com base no feedback deles.

Os resultados mostram que equipes voltadas a recursos não têm a melhor performance organizacional. Equipes assim parecem priorizar demais o desempenho de entrega, em detrimento da performance organizacional e do bem-estar dos funcionários da organização.

Equipes operacionais

» Equipes que focam no desempenho operacional podem trabalhar muito para otimizar métricas de sistema, como uso de CPU. Mas, caso não entendam o que um usuário espera de

um serviço, pode ser que elas recebam relatos frequentes de lentidão no funcionamento. Práticas de engenharia de confiabilidade do site (SRE), como identificar indicadores de nível de serviço que os usuários consideram importantes e definir objetivos de nível de serviço que sirvam para manter um usuário típico satisfeito, podem ajudar as equipes operacionais a adotar uma mentalidade mais centrada no usuário.

» Equipes de engenharia de plataforma

Equipes de engenharia de plataforma podem adotar uma abordagem “basta criar que os usuários chegarão até nós”.

Uma abordagem mais eficiente seria tratar os desenvolvedores como usuários da plataforma. Para que essa mudança de foco aconteça, as equipes de engenharia de plataforma precisam entender como os desenvolvedores trabalham hoje para conseguir identificar e eliminar áreas de atrito. As equipes podem usar as medidas de desempenho operacional e de entrega de software como indícios para monitorar se os esforços voltados à plataforma estão ajudando as equipes a melhorar os resultados.

» Líderes

Para ajudar a criar um ambiente que propicie o foco no usuário, os líderes criam uma estrutura de incentivos que recompense equipes que agregam valor aos usuários. Sem esse estrutura, pode ser que as equipes sintam que estão apenas medindo o número de recursos entregues ou a redução de interrupções do serviço. A DORA investigou o papel da liderança transformadora¹ e tem conselhos para líderes que estejam prontos para melhorar essa capacidade. Para mais informações, consulte “DevOps Capabilities: Transformational Leadership” em <https://dora.dev/devops-capabilities/cultural/transformational-leadership> (em inglês).



Recursos para começar

Para ter sucesso, é importante desenvolver capacidades mais focadas no usuário. Nossas descobertas de 2023 reforçam as de 2018,² quando notamos que capacidades de gerenciamento otimizado de produtos são preditoras do desempenho de entrega de software e da performance organizacional.

Para desenvolver o desempenho da sua equipe, adote capacidades com foco no usuário como feedback de clientes,³ visibilidade do trabalho no fluxo de valores,⁴ trabalho em lotes pequenos⁵ e experimentos em equipe.⁶

¹2017 State of DevOps Report. <https://dora.dev/publications/pdf/state-of-devops-2017.pdf>, 12-19

²2018 Accelerate: State of DevOps Report: Strategies for a New Economy. <https://dora.dev/publications/pdf/state-of-devops-2018.pdf>, 49-51

³“Customer feedback”. <https://dora.dev/devops-capabilities/process/customer-feedback/>

⁴“Visibility of work in the value stream”. <https://dora.dev/devops-capabilities/process/work-visibility-in-value-stream/>

⁵“Working in small batches”. <https://dora.dev/devops-capabilities/process/working-in-small-batches/>

⁶“Team experimentation”. <https://dora.dev/devops-capabilities/process/team-experimentation/>

Capítulo 3

Capacidades técnicas que predizem o desempenho

Pontos principais

É provável que o investimento em recursos e esforços de integração contínua, uma arquitetura com acoplamento flexível e a aceleração das revisões de código resultem em benefícios, por exemplo, melhorando o desempenho da equipe, da entrega de software e das operações. Isso acontece sem outros prejuízos e muitas vezes melhora o bem-estar das pessoas que trabalham no aplicativo ou serviço.

Introdução

No resumo executivo, explicamos as capacidades técnicas estudadas e como elas afetaram diferentes medidas de desempenho e bem-estar. Uma parte central da DORA sempre foi investigar e quantificar até que ponto diversos processos e capacidades técnicas predizem o desempenho.

Este ano, investigamos como as capacidades técnicas predizem o desempenho:

- Inteligência artificial
- Desenvolvimento baseado em linha principal
- Arquitetura com acoplamento flexível
- Integração contínua
- Revisão de código rápida

Analisamos como eles predizem estas medidas de desempenho:

- Desempenho da equipe
- Performance organizacional
- Desempenho de entrega de software
- Desempenho operacional

Além disso, testamos a ligação entre essas capacidades e diversos indicadores para determinar como eles afetam as pessoas que realizam o trabalho:

- Esgotamento
- Produtividade
- Satisfação no trabalho

Resultados

Os processos e capacidades técnicas estudados variaram, mas no geral têm efeitos positivos nas medidas principais de desempenho.

Processos e capacidades técnicas	Efeito no desempenho da equipe	Efeito na performance organizacional	Efeito no desempenho de entrega de software	Efeito no desempenho operacional				
IA		Nenhum efeito demonstrado		Aumento pequeno		Diminuição pequena		Diminuição substancial
Integração contínua		Aumento pequeno		Aumento pequeno		Aumento pequeno		Nenhum efeito demonstrado
Velocidade de revisão de código		Aumento pequeno		Nenhum efeito demonstrado		Aumento substancial		Aumento substancial
Arquitetura com acoplamento flexível		Aumento substancial		Aumento substancial		Aumento pequeno		Aumento substancial
Desenvolvimento baseado em linha principal		Aumento pequeno		Aumento pequeno		Aumento pequeno		Diminuição pequena

Equipes com acoplamento flexível, ou talvez equipes que têm uma arquitetura com acoplamento flexível, são capazes de alterar os sistemas de modo significativo sem envolver outras equipes. Assim, as equipes podem ser mais ágeis. Quando os especialistas do tema estão mais próximos da equipe, eles podem revisar o código com mais rapidez porque entendem melhor o impacto das mudanças. Um design com acoplamento flexível permite que a equipe teste, crie e implante sem que outras equipes acabem sendo um possível gargalo.

Mesmo com o pequeno impacto das nossas mudanças em uma arquitetura com acoplamento flexível, precisamos garantir que não estamos introduzindo conflitos com os outros desenvolvedores da equipe. Equipes que trabalham em lotes pequenos têm menos oportunidades de entrar em conflito. Isso garante que, a cada commit, o software seja criado e os testes automatizados sejam acionados, dando um feedback rápido aos desenvolvedores.

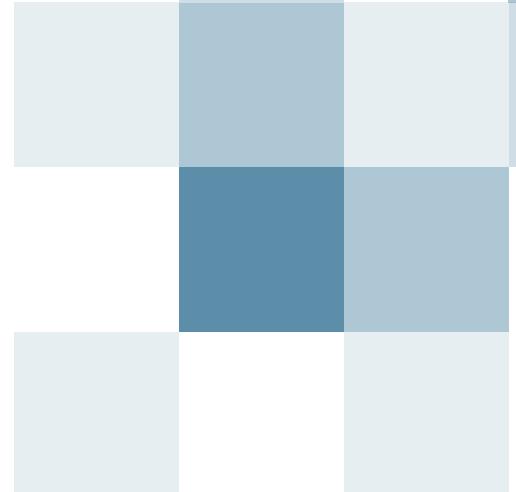
Equipes com revisões de código mais rápidas têm um desempenho de entrega de software 50% melhor. Processos eficientes de revisão de código geram melhorias no código, transferência de conhecimento, propriedade compartilhada sobre o código, propriedade sobre a equipe e transparência.

As revisões de código são gargalos para você? Avalie o processo de revisão de código e os efeitos dele no tempo de lead das mudanças para identificar oportunidades de aprimoramento. Pense nas seguintes questões:

- As revisões de código por pares estão integradas ao seu processo?
- Quanto tempo leva entre o preenchimento do código até a sua revisão?
- Qual é o tamanho médio dos lotes de revisões de código?
- Quantas equipes estão envolvidas nas revisões?
- How many different geographical locations are involved in your reviews?
- Sua equipe aprimora a automação de qualidade do código com base nas sugestões das revisões de código?

Quanto mais tempo passar entre o preenchimento e a revisão de código, maior o impacto negativo na eficiência do desenvolvedor e na qualidade do software entregue. O envolvimento de várias equipes em diversos locais geográficos estende a duração do processo, diminui o engajamento e aumenta os custos.¹

Em 2022, descobrimos que capacidades técnicas são complementares. Aumentar a velocidade da revisão de código pode ajudar a melhorar diversas capacidades técnicas, inclusive a facilidade de manutenção do código, a cultura de aprendizagem (transferência de conhecimento) e a criação de uma cultura generativa.²



Um dos benefícios de equipes com acoplamento flexível é a rapidez nas revisões de código, que melhora significativamente o desempenho operacional e o de entrega de software. Há várias formas de aumentar a eficiência das suas revisões de código. Se o código revisado só afetar o escopo da arquitetura da equipe, é mais provável que o revisor entenda melhor o impacto que o código vai ter no sistema. Quanto menor a revisão de código, mais fácil é para o revisor entender as implicações da mudança. Trabalhar em pequenos lotes é uma maneira de melhorar o ciclo de feedback, a eficiência e o foco da equipe.³ A programação em pares é uma prática que pode reduzir o tempo de revisão de código independentemente da arquitetura e das práticas de integração atuais.⁴

Além disso, essas capacidades e processos não têm um impacto negativo no bem-estar das pessoas que realizam o trabalho. Na verdade, a maioria desses elementos são preditores de melhorias no bem-estar desses indivíduos.

¹ Investigating the effectiveness of peer code review in distributed software development based on objective and subjective data. <https://jserd.springeropen.com/articles/10.1186/s40411-018-0058-0>

² Expectations, Outcomes, and Challenges of Modern Code Review. <https://dl.acm.org/doi/10.5555/2486788.2486882>

³ “Working in small batches”. <https://dora.dev/devops-capabilities/process/working-in-small-batches/>

⁴ “On Pair Programming”. <https://martinfowler.com/articles/on-pair-programming.html>

Processos e capacidades técnicas	Efeito no esgotamento*	Efeito na satisfação no trabalho	Efeito na produtividade
IA	↓ Diminuição pequena	↑ Aumento pequeno	↑ Aumento pequeno
Integração contínua	⊖ Nenhum efeito	↑ Aumento pequeno	⊖ Nenhum efeito
Velocidade de revisão de código	↔ Diminuição substancial	↑ Aumento pequeno	↑ Aumento pequeno
Arquitetura com acoplamento flexível	↔ Diminuição substancial	↑↑ Aumento substancial	↑↑ Aumento substancial
Desenvolvimento baseado em linha principal	↑↑ Aumento substancial	⊖ Nenhum efeito	⊖ Nenhum efeito

*Note que o esquema de cores está trocado para o esgotamento. Isso foi feito porque diminuir o esgotamento é positivo!

Notamos que o uso de uma arquitetura com acoplamento flexível, integrações contínuas e revisões de código eficientes permite que as equipes melhorem os resultados organizacionais mantendo e às vezes aprimorando o bem-estar.

Quando as equipes têm autonomia para aprimorar e manter um sistema confiável que oferece valor aos usuários, elas sentem mais satisfação no trabalho e aumentam o desempenho da equipe e da entrega de software. A arquitetura tem um papel significativo na capacidade da equipe de focar no usuário e aprimorar a entrega de software. Começando pequeno e focando no usuário, as equipes conseguem melhorar significativamente o desenvolvimento baseado em linha principal, arquitetura com acoplamento flexível, integração e entrega contínuas e SRE. Para aprimorar sua capacidade técnica, encontre oportunidades para as equipes fazerem experimentos e estarem sempre melhorando.⁵



⁵ "Team experimentation". <https://dora.dev/devops-capabilities/process/team-experimentation>

Benefícios de entrega contínua

Autor: Dave Farley

O princípio fundamental da entrega contínua (CD) é trabalhar para que o software esteja sempre pronto para ser lançado. Para isso, o trabalho precisa ter alta qualidade. Assim, fica mais fácil corrigir os problemas detectados e preparar o software para ser lançado mais rapidamente.

Para que o software esteja sempre pronto para ser lançado, precisamos trabalhar para estabelecer um ciclo rápido de feedback e recuperar rapidamente em caso de falhas.

Como leitor do relatório deste ano, imagino que essas ideias soem familiares. As métricas de **Estabilidade** (taxa de falha nas mudanças e tempo de recuperação após falha ao implantar) são relacionadas à qualidade, e as métricas de **Capacidade de processamento** (tempo de lead das mudanças e frequência de implantação) envolvem feedback e facilidade para detectar qualquer problema.

Se você pratica a CD, tem uma pontuação alta de **Estabilidade e capacidade de processamento**. Se você tem uma pontuação alta de **Estabilidade e capacidade de processamento**, é difícil imaginar que você não esteja praticando a CD para alcançar essa pontuação.

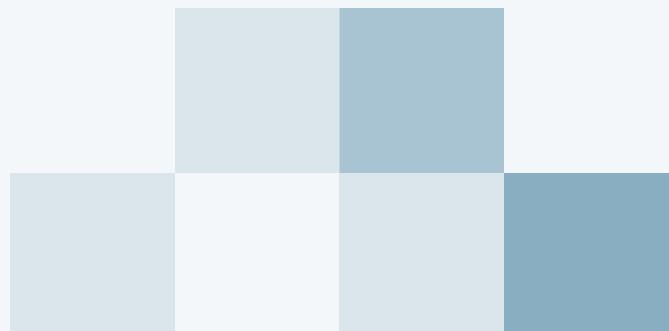
A análise deste ano também aborda formas de aprimorar as capacidades que impulsionam o desempenho. A CD, a capacidade de lançar alterações de todos os tipos sob demanda de maneira rápida, segura e sustentável, está muito ligada a várias capacidades técnicas. Em outras

palavras, essas capacidades funcionam porque criam um ambiente que viabiliza a CD. A prática da CD, por sua vez, serve como mecanismo para essas capacidades predizerem um bom desempenho de entrega de software.

Preparar o software para ser lançado é um padrão importante no desenvolvimento, por isso é um dos focos da CD. Isso importa porque é um atestado de qualidade subjetivo, porém definido e atrelado ao contexto. O nível de rigor que define se o software está pronto para o lançamento pode ser diferente se estivermos trabalhando em sistemas críticos para a segurança ou em um software para uma loja de bolos. Mas, nos dois casos, isso define que fizemos tudo necessário para afirmar que o código está pronto, bom o suficiente e seguro o bastante para ser lançado para os usuários.

Por isso, otimizar com foco no pregar das mudanças para o lançamento também significa otimizar pensando na qualidade minimamente aceitável para nosso sistema nesse contexto.

Equipes que priorizam buscar e agir com base em feedback ágil têm um melhor desempenho de entrega de software.



Para mim, é surpreendente que a integração contínua (CI) e o desenvolvimento baseado em linha principal não tenham um impacto maior no desempenho de entrega de software. A CI em particular parece ser bastante fundamental, então isso é um tanto desafiador para minha visão de mundo, mas desconstruir isso é complicado. Por exemplo, como podemos ter uma pontuação alta para capacidade de processamento se o código não é integrado ou como podemos garantir que a estabilidade é alta sem verificar isso? Para mim, isso é feito na CI, que é um importante mediador do desempenho de entrega de software. Isso é um problema de interpretação ou algo mais profundo e importante? É algo para se pensar!

Processos e capacidades técnicas	Efeito no desempenho de entrega de software	Mediado pela entrega contínua?*
IA	 Diminuição pequena	<input checked="" type="checkbox"/> Não
Integração contínua	 Aumento pequeno	<input checked="" type="checkbox"/> Sim, totalmente
Velocidade de revisão de código	 Aumento substancial	<input checked="" type="checkbox"/> Sim, parcialmente
Arquitetura com acoplamento flexível	 Aumento pequeno	<input checked="" type="checkbox"/> Sim, parcialmente
Desenvolvimento baseado em linha principal	 Aumento pequeno	<input checked="" type="checkbox"/> Sim, totalmente

*A mediação é um teste que avalia possíveis mecanismos ou caminhos relacionados a um efeito. É possível dizer, por exemplo, que “os dados apoiam a hipótese de que o efeito do desenvolvimento baseado em linha principal no desempenho de entrega de software ocorre por meio da implantação contínua (o mediador)”. A mediação completa ocorre quando existe a percepção de que todo o efeito pode ser explicado pelo mediador. A mediação parcial ocorre quando apenas parte do efeito pode ser explicada pelo mediador.

Otimização de capacidades e processos organizacionais

Sabemos que a cultura impulsiona o sucesso. Mas o que impulsiona a cultura? Essa é uma pergunta interessante, e a resposta que todos adoram dar é: depende!

Da perspectiva do profissional, o aprimoramento do trabalho diário causa um impacto positivo em elementos culturais, como compartilhamento de riscos, aumento da cooperação e segurança psicológica. Por exemplo, integrar mudanças regulares na ramificação principal do sistema de controle de versões aumenta o compartilhamento de conhecimentos e a colaboração. Quando as equipes de segurança colaboram com os desenvolvedores na política como código, isso aumenta a confiança entre as equipes e em relação às mudanças implantadas.

Da perspectiva da liderança, a primeira etapa é desenvolver a consciência e realizar treinamentos sobre a importância da cultura. A liderança transformacional⁶ pode ajudar a criar um ambiente que não aponta culpados e que incentiva a experimentação e a aprendizagem, além de confiar e dar voz aos profissionais. Engenheiros estão lá para resolver problemas complexos, não só responder a solicitações de tarefas. Para isso, eles precisam ter visibilidade sobre o negócio e autonomia para agir. Em última instância, a cultura é fruto da liderança.

Em um cenário ideal, os melhores resultados surgem quando a cultura é analisada por quem está embaixo e por quem está em cima na hierarquia.

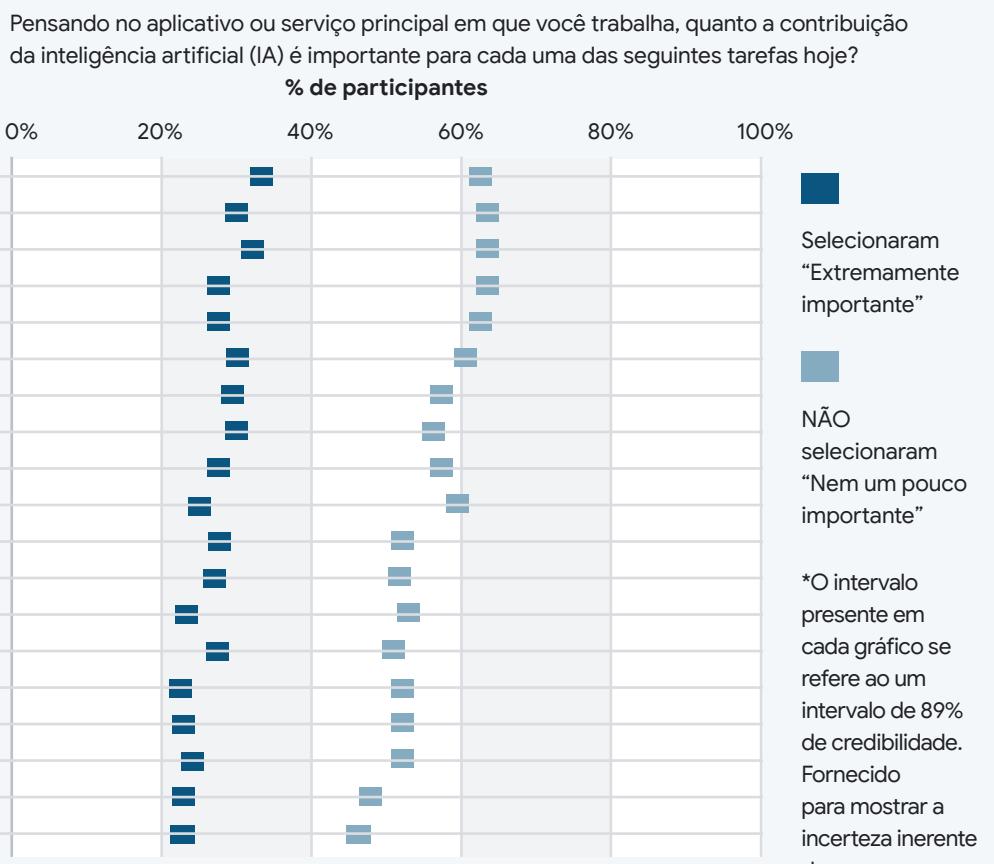
⁶ “Transformational leadership”. <https://dora.dev/devops-capabilities/cultural/transformational-leadership>

Inteligência artificial (IA)

Alguns analistas e tecnologias preveem que a IA vai aumentar o desempenho das equipes de software sem ter um impacto negativo no bem-estar profissional das pessoas. Até agora, nossa pesquisa não confirma isso. Nossas evidências sugerem que a IA melhora medidas individuais de bem-estar (como esgotamento e satisfação no trabalho), mas tem um efeito neutro ou talvez negativo em aspectos do grupo (como o desempenho da equipe ou da entrega de software).

Nossa hipótese é que a etapa inicial da adoção de ferramentas de IA nas empresas pode ajudar a explicar esse misto de evidências. É provável que empresas de grande porte estejam testando diferentes ferramentas de IA antes de decidir se vão usá-las de modo mais abrangente. Há muito otimismo sobre o potencial de ferramentas de desenvolvimento com IA, conforme demonstrado pela maioria das pessoas que incorporaram ao menos um pouco de IA às tarefas mencionadas na pesquisa. Isso é mostrado no gráfico abaixo. Acreditamos que vai levar algum tempo até as ferramentas com tecnologia de IA serem usadas de modo amplo e coordenado no setor.

Importância da IA Contribuição da IA em tarefas técnicas



Temos muito interesse em observar como a adoção aumenta ao longo do tempo e o impacto disso em medidas de desempenho e resultados importantes para as organizações.

O papel fundamental da documentação



Pontos principais

É fundamental ter uma documentação de boa qualidade. Ela ajuda a implementar as capacidades técnicas com êxito e intensifica o impacto dessas capacidades na performance organizacional. A documentação também tem um impacto positivo nos resultados, como desempenho da equipe, produtividade e satisfação no trabalho. No entanto, ter mais documentação de qualidade não garante um bem-estar melhor para todos: conforme a qualidade da documentação aumenta, alguns participantes relatam níveis mais altos de esgotamento.

Introdução

Este ano, analisamos com mais profundidade a documentação interna, ou seja, o conhecimento escrito que as pessoas da organização usam no dia a dia. Investigamos o impacto da documentação nas capacidades técnicas e nos resultados principais.

Para avaliar a qualidade da documentação, medimos quanto ela é confiável, encontrável, atualizada e relevante. Em seguida, calculamos uma pontuação para toda a experiência de documentação. Não estamos avaliando cada página da documentação, mas sim ela como um todo.

Resultados

A documentação é fundamental porque eleva e intensifica capacidades técnicas

Conforme descobrimos em 2021¹ e 2022², a qualidade da documentação continua elevando o sucesso da implementação de todas as capacidades técnicas estudadas.

Como mostrado na tabela a seguir, a qualidade da documentação também intensifica o impacto de cada capacidade técnica na performance organizacional, similar ao que constatamos em 2022³.

Capacidade técnica	Intensificação do impacto na performance organizacional*
Integração contínua	2,4x*
Entrega contínua	2,7x*
Desenvolvimento baseado em linha principal	12,8x*
Arquitetura com acoplamento flexível	1,2x*
Práticas de confiabilidade	1,4x*
Contribuição da inteligência artificial	1,5x*

*Cálculo:

Impact of technical capability with high documentation quality

Impact of technical capability with low documentation quality

¹ Accelerate State of DevOps 2021. <https://dora.dev/publications/pdf/state-of-devops-2021.pdf>

² 2022 Accelerate State of DevOps Report. <https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report.pdf>

³ 2022 State of DevOps Report data deep dive: Documentation is like sunshine. <https://cloud.google.com/blog/products/devops-sre深深洞见2022年DevOps报告中的文档>

A documentação prediz a presença de pessoas e organizações produtivas e felizes

Além de melhorar as capacidades técnicas, descobrimos que a documentação de qualidade tem um impacto positivo no bem-estar das pessoas: ela reduz o esgotamento, melhora a satisfação no trabalho e aumenta a produtividade. Descobrimos que parte desse impacto se deve ao aumento no compartilhamento de conhecimento que a documentação de qualidade viabiliza.

Esse efeito não é tão surpreendente. Tudo fica mais fácil quando o conhecimento é compartilhado, porque você sabe como cada tarefa precisa ser feita e o trabalho fica menos frustrante.

Aspectos do bem-estar	Efeito de uma documentação de qualidade
Esgotamento	 Diminuição substancial
Satisfação no trabalho	 Aumento substancial
Produtividade	 Aumento substancial

Uma documentação de qualidade também eleva os resultados principais, com efeitos no desempenho da equipe, na performance organizacional e no desempenho operacional.

Resultados principais	Efeito de uma documentação de qualidade
Desempenho da equipe	 Aumento substancial
Performance organizacional	 Aumento substancial
Desempenho de entrega de software	 Nenhum efeito*
Desempenho operacional	 Aumento substancial

* Continuamos surpresos com o fato de que uma documentação de qualidade não afeta o desempenho de entrega de software. Este é o segundo ano em que constatamos esse comportamento.

O que está por trás desse efeito positivo nos três resultados principais? Como leitor, é útil ter uma documentação clara. O processo de escrita também pode ser um fator relevante. Para criar uma documentação de alta qualidade, é necessário que antes as equipes determinem os processos. A documentação pode ajudar as equipes da organização a discutirem abertamente e concordarem sobre o que precisa ser feito e como.

A documentação de qualidade também funciona como um repositório do conhecimento da equipe, mesmo que os membros mudem. Ela ajuda a ampliar o conhecimento na organização e ao longo do tempo.

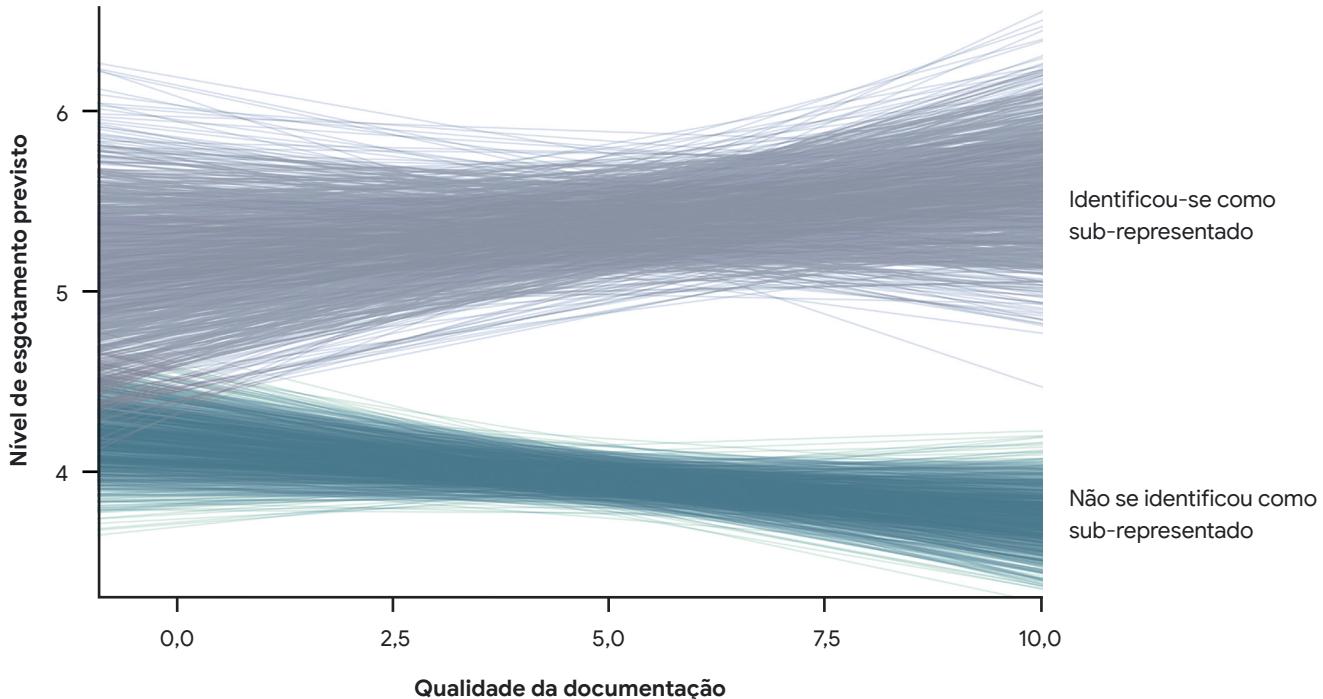
A documentação está ligada a uma redução do bem-estar de algumas pessoas?

Notamos uma tendência inesperada nas respostas de participantes que se identificam como sub-representados. Para esse grupo, a qualidade da documentação está vinculada a um aumento do esgotamento.

Para esta descoberta, também analisamos o gênero, mas surpreendentemente esse fator não teve efeito. A documentação de alta qualidade reduziu significativamente o esgotamento de participantes que se identificaram como do gênero masculino, feminino ou autodeterminado. No entanto, independente da identidade de gênero, as pessoas que se identificaram como sub-representadas notaram uma taxa maior de esgotamento quando têm uma documentação de qualidade.

O gráfico a seguir mostra previsões simuladas baseadas nos dados. No conjunto mais baixo, temos uma redução no esgotamento para a maioria dos participantes conforme a qualidade da documentação aumenta. No entanto, no conjunto mais alto, o esgotamento aumenta de modo significativo para pessoas que se identificam como sub-representadas.

Este gráfico mostra mil linhas simuladas para cada grupo. Linhas mais densas apontam que a inclinação é mais provável com base em nossos dados.



Essa descoberta é similar para qualidade da documentação, cultura generativa e estabilidade da equipe: conforme esses atributos aumentam, o esgotamento também aumenta para pessoas que se identificam como sub-representadas. O que está acontecendo no caso da documentação?

Elaborar e manter uma documentação de alta qualidade exige esforço. É um trabalho técnico, com impacto significativo nas capacidades técnicas, produtividade da equipe e performance organizacional. Também é possível que ele não seja reconhecido, apesar da importância e impacto que tem. Será que as pessoas que se identificam como sub-representadas estão realizando um volume desproporcional desse trabalho e, se estiverem, isso ajuda a explicar o efeito no esgotamento?

Será que depender da documentação é um problema? Melhorar a qualidade da documentação também aumenta o compartilhamento de conhecimentos para alguns participantes? Ou, se houver esse aumento, ele é suficiente para contrapor outros aspectos que causam esgotamento na equipe?

É possível que haja outro fator que melhora a qualidade da documentação, mas gera ou perpetua o esgotamento para participantes que se identificam como sub-representados. São necessárias mais pesquisas.

Há indícios de que o cargo dentro da equipe importa. Aspectos do espaço de trabalho (como uma documentação de qualidade) oferecem benefícios significativos para a equipe e a organização no geral. No entanto, pode ser que eles estejam relacionados a impactos negativos para alguns indivíduos. Vamos investigar isso mais a fundo no [Capítulo 8: A importância de "como", "quando", "por quê" e "quem"](#).

⁴ Accelerate State of DevOps 2021, 22. <https://dora.dev/publications/pdf/state-of-devops-2021.pdf#page=22>

Recursos para começar

Confira o relatório de 2021 para conhecer práticas que melhoraram a qualidade da documentação.⁴ Este ano, também descobrimos que a distribuição do trabalho, inclusive processos formais de distribuição do trabalho relacionado à documentação, aumenta significativamente a qualidade da documentação.

Há vários recursos e treinamentos sobre escrita técnica. Aqui estão algumas recomendações em inglês para você saber mais:

- Sociedade de comunicação técnica (stc.org)
- Cursos de escrita técnica para engenheiros (developers.google.com/tech-writing)
- Elaboração de documentos (writethedocs.org)

A confiabilidade aumenta o desempenho

Pontos principais

Boas práticas de confiabilidade são indícios de bom desempenho operacional, desempenho da equipe e performance organizacional. Os dados mostram que a evolução dessas práticas segue um caminho não linear, ou seja, pode haver momentos em que a performance parece estagnada, apesar de melhorias nas capacidades da organização. No entanto, o comprometimento com essas práticas continua sendo um preditor de bons resultados.

Introdução

A confiabilidade é um termo amplamente usado no espaço de operações de TI. A confiabilidade define até que ponto um serviço cumpre as metas a que se propõe com base em aspectos como disponibilidade, desempenho e correção. Uma abordagem comum para melhorar a confiabilidade é a SRE, que foi criada no Google (<https://sre.google>) e agora é usada em várias organizações. A SRE prioriza aprendizado empírico, colaboração multifuncional, amplo uso de automação e técnicas de medição como objetivos de nível de serviço (SLOs).

Várias organizações usam práticas de confiabilidade sem chamá-las de SRE. Outros termos incluem engenharia da produção, equipes de plataforma, equipes de infraestrutura, TechOps, entre outros. Para avaliar essas práticas da forma mais objetiva possível, tivemos o cuidado de usar uma linguagem neutra e descritiva no texto da pesquisa apresentado aos participantes.

Também coletamos dados sobre os resultados da engenharia de confiabilidade: até que ponto as equipes conseguem cumprir as metas relacionadas a esse aspecto. Ao lado de outras capacidades, nosso modelo preditivo reflete práticas e resultados de confiabilidade (que chamamos de **desempenho operacional**).

Práticas de confiabilidade

Pedimos para os participantes pensarem sobre a confiabilidade considerando três aspectos operacionais essenciais. Primeiro, existem planos de mitigação para as dependências? Segundo, os planos de recuperação de desastres são testados regularmente por meio de simulações de interrupção, failovers práticos ou exercícios de decisão? Por fim, quando não alcançam as metas de confiabilidade, eles implementam melhorias ou redefinem prioridades e ajustam o trabalho?

Acreditamos que essas medidas representam o espírito de uma equipe que segue princípios estabelecidos de SRE como “assumir riscos” e “medir a felicidade do usuário”. Uma equipe assim define uma meta razoável alinhada à felicidade do usuário. Em seguida, ela realiza testes para garantir que vai conseguir atingir a meta e muda os planos caso tenha dificuldades. Usamos isso como indicador de uma equipe que consegue “fazer SRE” sem vincular as avaliações da equipe a implementações específicas de SRE.

Resultados

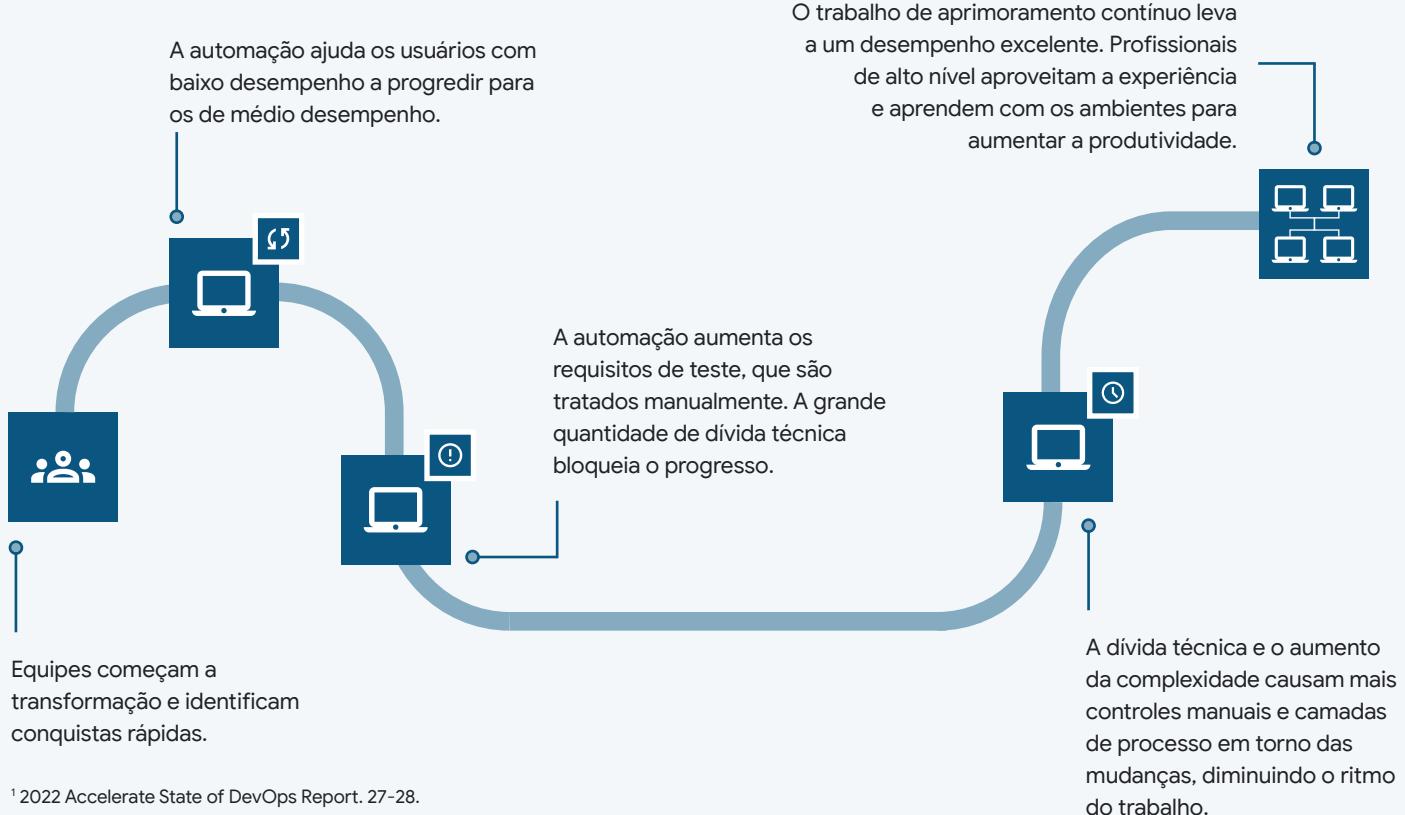
Como confirmar a curva J de práticas de confiabilidade

Desde 2018, a DORA elaborou a teoria de que há uma relação não linear (Figura 1) entre desempenho operacional e práticas como a automação. Conforme

aprofundamos nossas investigações sobre práticas de confiabilidade, temos notado evidências desse padrão nos dados da pesquisa.

Em 2022, medimos isso de forma direta. Nós perguntamos às equipes e observamos que a relação entre práticas e resultados de confiabilidade de fato condiz com esse tipo de curva não linear (Figura 2). Isso sugere que as equipes tiveram um aumento significativo na confiabilidade apenas depois de adotar várias práticas nesse sentido.¹ No entanto, esses dados não pareciam mostrar o contexto geral. A curva de 2022 passou a impressão de que a SRE só serve para especialistas e não vale o esforço, o que vai contra a experiência de várias equipes de SRE. Precisamos de mais dados.

Figura 1: curva J hipotética de 2018



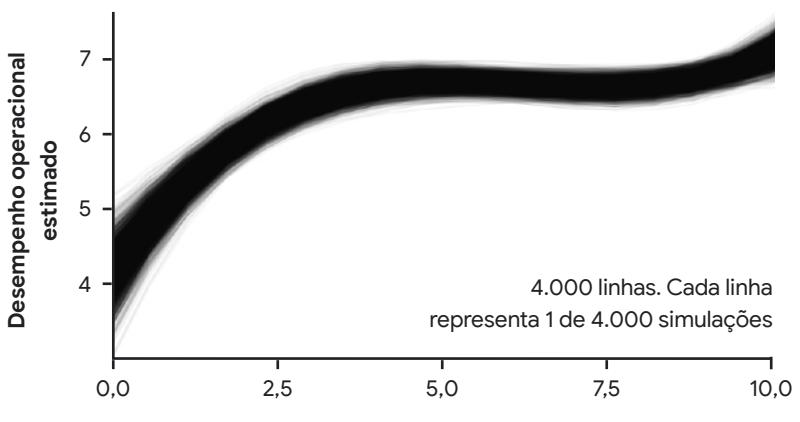
¹ 2022 Accelerate State of DevOps Report. 27-28.

https://dora.dev/research/2022/dora-report/_2022-dora-accelerate-state-of-devops-report.pdf#page=27

Figura 2: curva de 2022



Figura 3: curva de 2023



Em 2023, conseguimos fazer mais perguntas para definir uma curva que corresponde melhor às nossas experiências. A nova é mais próxima da curva J hipotética de transformação descrita no relatório de 2018 (consulte a seção [Metodologia](#) para saber mais e conferir como realizamos nossa análise). Isso sugere que adotar práticas de confiabilidade de fato gera conquistas iniciais, que são seguidas por um período de constância em meio aos novos desafios causados pelo aumento da complexidade, para por fim gerar outro aumento no desempenho operacional. Os resultados reforçam o que constatamos em várias equipes.

Essa curva é importante por alguns motivos:

- Ela ajuda as empresas a justificar e financiar a adoção inicial da SRE, mesmo que não estejam em busca de níveis extremos de confiabilidade ou não tenham a intenção de fazer investimentos significativos na SRE. Adotar até pequenas práticas de confiabilidade pode melhorar o desempenho operacional, gerando ainda mais benefícios para o desempenho da equipe e a performance organizacional.

- Ela prepara as empresas que querem fazer grandes investimentos em confiabilidade para o período de constância. É tentador esperar que investimentos a longo prazo em SRE gerem resultados positivos lineares, mas dados apontam que não é assim que isso funciona. Se as equipes estiverem cientes dessa não linearidade da curva, elas podem decidir se vão fazer esse investimento e planejar com antecedência para garantir que essa iniciativa não seja abandonada antes de alcançar todos os benefícios.
- Mudanças como essas exigem uma transformação cultural.² Nós descobrimos que o sucesso é resultado de uma combinação de mudanças de baixo para cima e de cima para baixo. As equipes podem adotar práticas de confiabilidade e aproveitar os frutos imediatos. Depois, a liderança pode mostrar esses benefícios para outras equipes com o intuito de reforçar e incentivar iniciativas desse tipo. Esses incentivos e programas estruturados podem ser elaborados com base na curva J.

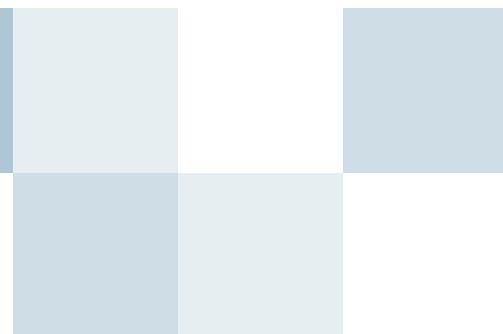
² “How to transform” - <https://dora.dev/devops-capabilities/cultural/devops-culture-transform/>

Práticas de confiabilidade e bem-estar

Práticas operacionais tradicionais são altamente reativas e frequentemente mais preocupadas com a integridade do sistema técnico, não com a felicidade dos usuários. Alertas imediatos para coisas que não afetam as experiências dos usuários, tarefas manuais repetitivas, receio de cometer erros e experiências similares levam ao esgotamento e prejudicam o bem-estar das pessoas na equipe.

Constatamos o oposto em equipes com práticas de confiabilidade. Essas equipes relatam níveis maiores de produtividade e satisfação no trabalho e níveis menores de esgotamento do que as que não seguem essas práticas. Nossa suspeita é que esse aumento no bem-estar seja impulsionado por práticas de SRE já publicadas:

- Redução de tarefas repetitivas³
- Análises post-mortem sem apontar culpados⁴
- Autonomia da equipe⁵
- Escalonamento sublinear de equipes⁶



³Beyer, Betsy, et al. Site Reliability Engineering: How Google Runs Production Systems (O'Reilly, 2016), 49–54. <https://sre.google/sre-book/eliminating-toil/>

⁴Ibid, <https://sre.google/sre-book/postmortem-culture/>

⁵Beyer, Betsy, et al. The Site Reliability Workbook (O'Reilly, 2018), <https://sre.google/workbook/team-lifecycles/>

⁶Brookbank, James, and McGhee, Steve. Enterprise Roadmap to SRE (O'Reilly, 2022), 11. <https://sre.google/resources/practices-and-processes/enterprise-roadmap-to-sre/>

Desempenho operacional

Também pedimos para os participantes descreverem o desempenho operacional do serviço deles. Primeiro, perguntamos com que frequência eles recebem feedback direto dos usuários sobre problemas na confiabilidade do serviço. Depois, perguntamos com que frequência o serviço fica indisponível, lento ou tem erros operacionais.

Práticas de confiabilidade intensificam o desempenho da equipe e a performance organizacional, por meio do desempenho operacional

Quando adotam práticas de confiabilidade, o desempenho operacional das equipes melhora. Quando uma organização consegue operar a frota de produção com eficiência, os outros resultados são *intensificados*. Se os resultados já estiverem altos, as práticas de confiabilidade fazem com que eles aumentem ainda mais. Se os resultados estiverem baixos, as práticas de confiabilidade não vão ter nenhum efeito.

Mesmo assim, os sistemas confiáveis ainda precisam ter as capacidades de software ideais e entregá-las para os clientes de maneira eficiente. Isso faz sentido porque a SRE não foi criada para funcionar no vácuo. Atingir as metas de confiabilidade é uma métrica importante para o sucesso das equipes de SRE, e o reflexo disso é o desempenho operacional. Mesmo sendo provável que práticas de confiabilidade tenham outros benefícios, os dados sugerem que o mais essencial é o impacto no desempenho operacional. Além disso, o aumento do desempenho operacional tem outros benefícios além da integridade do serviço. Na verdade, há evidências de que o uso de práticas de confiabilidade estão ligadas a um aumento no bem-estar dos profissionais.

O desempenho operacional afeta o bem-estar

Uma percepção comum no setor é que serviços altamente confiáveis têm um impacto negativo no bem-estar dos operadores de serviço, por exemplo, por causa de atividades imediatas ou manutenções de emergência fora do horário útil.

No entanto, descobrimos que o alto desempenho operacional resulta em menos esgotamento, maior produtividade e mais satisfação no trabalho. Isso está de acordo com o princípio da SRE de reduzir tarefas repetitivas;⁷ a automatização de aspectos manuais das operações aumenta a satisfação e reduz o ônus da equipe.

A performance organizacional e o desempenho da equipe intensifica o desempenho operacional

Descobrimos que o desempenho operacional tem um impacto positivo substancial no desempenho da equipe e na performance organizacional. Isso não deve ser surpresa para quem segue o movimento de DevOps. Quando a equipe sabe operar uma máquina com eficiência, ela consegue conquistar mais e fazer a organização ter mais sucesso.

O desempenho operacional intensifica o desempenho de entrega de software

Embora o desempenho de entrega de software possa melhorar o desempenho da equipe e a performance organizacional, o desempenho operacional tem um impacto mais significativo nesses dois aspectos. Além disso, equipes com bom desempenho de entrega de software não conseguem alcançar altos níveis de desempenho da equipe e de performance organizacional sem também atingir um alto desempenho operacional. É necessário atingir os dois. Na verdade, equipes que melhoraram o desempenho de entrega de software sem atingir níveis similares de desempenho operacional acabam tendo resultados organizacionais menores. Então, se você consegue escrever um software incrível em pouco tempo, mas não é capaz de executar a produção de modo que ela atenda às expectativas do público-alvo, o mercado não vai recompensar seu trabalho.

O que está faltando e o que vem a seguir?

Acreditamos que mais medidas podem nos ajudar a entender essas interações. Por exemplo, uma pergunta comum neste ano foi como o gerenciamento de custos afeta essas capacidades e resultados.

Algumas organizações têm mais limitações financeiras, e isso influencia como elas planejam e tomam decisões. Da mesma forma, temos a teoria de que práticas de confiabilidade podem emergir de culturas altamente colaborativas,⁸ mesmo sem serem buscadas ou planejadas de modo explícito. Queremos entender melhor como as equipes evoluem as operações de TI atuais e como isso afeta a confiabilidade do sistema, o desempenho da equipe e o bem-estar.

Acima de tudo, queremos ouvir você. Junte-se a nós e outros profissionais na [comunidade da DORA](#).⁹ A SRE ainda é um campo novo. O impacto dela é diferente para cada organização que adota práticas de confiabilidade ou que percebe que já segue a SRE há muito tempo. Essas mudanças são lentas e precisam ser medidas de forma consistente para mostrar o avanço ao longo do tempo. Enquanto comunidade, podemos compartilhar o que funciona e ajudar uns aos outros nessa jornada.

⁷ Beyer, Betsy, et al. Site Reliability Engineering: How Google Runs Production Systems (O'Reilly, 2016), 49–54. <https://sre.google/sre-book/eliminating-toil/>

⁸ Brookbank, James, and McGhee, Steve. Enterprise Roadmap to SRE (O'Reilly, 2022), 5. <https://sre.google/resources/practices-and-processes/enterprise-roadmap-to-sre/>

⁹ DORA Community, <https://dora.community/>

Como é a SRE no Google

No Google, a SRE evoluiu ao longo de duas décadas dentro de uma organização em ascensão e com bom financiamento. O poder da SRE foi necessário para viabilizar o imenso crescimento da Pesquisa Google e do Google Ads sem drenar todos os recursos financeiros. Logo no início, como esses produtos precisam reagir em tempo real, foi detectada a necessidade de garantir uma alta confiabilidade. Em outras palavras, um erro passageiro na Pesquisa Google ou no Google Ads pode imediatamente resultar na perda do cliente sem a chance de tentar de novo. Um anúncio dinâmico precisa ser calculado em milissegundos, e uma pesquisa lenta prejudica toda a promessa de marca do Google.

Ao mesmo tempo, a SRE estava sendo desenvolvida em outro tipo de empresa: uma organização de baixo para cima, baseada em engenharia, que preferia criar em vez de comprar. O Google usou uma equipe de engenheiros de confiabilidade do site (SREs) com bom desempenho acadêmico em disciplinas de ciência da computação como sistemas distribuídos e design de compilador. Esse DNA cultural serviu como um ambiente rico para o surgimento e a ascensão da SRE. A equipe de SRE recebeu não só os elementos essenciais da produção, mas também encontrou maneiras novas e inovadoras de escalonar sistemas.

Mas como a SRE foi ampliada ao longo do tempo? Três palavras: de forma sublinear. Ou seja, o Google não podia dobrar o número de SREs toda vez que dobrava de tamanho. Devido ao ritmo com que os primeiros produtos do Google estavam sendo expandidos para atender à demanda global (e a introdução de novos produtos, como Gmail, Google Maps, Android, YouTube e Google Cloud), não foi possível escalar essas SREs novas e escassas para acompanhar o crescimento da base de clientes.

¹⁰ Benjamin Treynor Sloss, vice-presidente de engenharia. Site Reliability Engineering: How Google Runs Production Systems (O'Reilly, 2016).

<https://sre.google/sre-book/part-i-introduction/>, <https://sre.google/sre-book/introduction/#id-2opuzSjFr>

Isso levou à criação de um lema:

- A SRE não deve acompanhar linearmente o aumento no número de usuários.
- A SRE não deve acompanhar linearmente o aumento no número de servidores.
- A SRE não deve acompanhar linearmente o aumento no número de clusters.
- A SRE não deve acompanhar linearmente o aumento no número de serviços.

Assim, foi desenvolvida uma estrutura de gerenciamento para manter esse modelo de crescimento restrito. As SREs se alinharam e cooperaram entre si sem deixar de lado as equipes de desenvolvimento de produto com que elas estavam trabalhando. As SREs reportavam a uma cadeia de gerenciamento própria, liderada por Ben Treynor-Sloss¹⁰. Essas equipes se dividiram em áreas de produto (PAs) que eram alinhadas diretamente a PAs de desenvolvimento de produto (“dev”). As equipes das SREs trabalharam com as equipes de dev para decidir a melhor maneira de usar as SREs disponíveis. As equipes de dev podiam financiar novas SREs diretamente. Nem todas as equipes de dev criaram produtos voltados ao usuário. Muitas delas eram equipes de infraestrutura

compartilhada, como o Bigtable¹¹ (armazenamento de dados estruturados), o Borg¹² (programação de computação) e o Colossus¹³ (armazenamento distribuído). Esses serviços de infraestrutura compartilhada permitiam que as equipes voltadas a clientes escalonassem sem ter uma equipe de SRE dedicada.

Mantendo essas equipes na própria organização, as equipes conseguiam adotar um processo consistente de contratação e promoção. Era comum que as equipes de SRE fossem dez ou mais vezes menores do que as equipes de dev. Por isso, era importante garantir que as equipes de SRE tivessem autonomia e não fossem contra os princípios de SRE.

As SREs desenvolveram produtos próprios, inclusive com gerentes de produto internos. Os clientes dos produtos de SRE eram outras equipes que queriam encontrar formas melhores de executar a produção. As equipes de SRE desenvolveram produtos centrados em resposta a incidentes, monitoramento e observabilidade, gerenciamento de lançamento, planejamento de capacidade e resolução de problemas.

O processo de crescimento das equipes de SRE sempre levou em consideração a sustentabilidade e a integridade da equipe. Por exemplo, uma nova equipe de SRE precisava ter no mínimo 12 pessoas: 2 locais com 6 membros. Isso permitia cobrir todos os fuso horários e garantir um bom equilíbrio entre trabalho e vida pessoal para todos os membros da equipe, sem gerar esgotamento.

A SRE continua se adaptando até hoje. Nem todas as equipes seguem os modelos descritos acima, e algumas optaram por caminhos alternativos. Organizações grandes, como a SRE do Google, também são afetadas pelas condições de mercado, como qualquer outro grupo grande. Por isso, a flexibilidade é importante. Mas, acima de tudo, as equipes de SRE precisam seguir seus princípios: assumir riscos, medir níveis de serviço, eliminar tarefas repetitivas, abraçar a automação e prezar pela simplicidade.



¹¹ Chang, Fay, et al. Bigtable: A Distributed Storage System for Structured Data, 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI), {USENIX} (2006), pp. 205-218, <https://research.google/pubs/pub27898/>

¹² Verma, Abhishek, et al. Large-scale cluster management at Google with Borg, Proceedings of the European Conference on Computer Systems (EuroSys), ACM, Bordeaux, France (2015), <https://research.google/pubs/pub43438/>

¹³ Hildebrand, Dean, et al. Colossus under the hood: a peek into Google's scalable storage system. 19 de abril de 2021 - <https://cloud.google.com/blog/products/storage-data-transfer/a-peek-behind-colossus-googles-file-system>

Capítulo 6

A infraestrutura flexível é a chave do sucesso

Pontos principais

A infraestrutura flexível é um preditor do desempenho de equipe, da performance organizacional, do desempenho operacional e do desempenho de entrega de software. A computação em nuvem é um dos principais elementos que viabiliza a infraestrutura flexível, mas esse benefício não é automático: nossos dados mostram que a *maneira* de usar a nuvem também importa.

Introdução

Ao longo de boa parte da pesquisa DORA, perguntamos aos profissionais sobre a infraestrutura, principalmente características essenciais da computação em nuvem, conforme a definição do Instituto Nacional de Padrões e Tecnologia dos EUA (NIST):¹

- Autoatendimento sob demanda
- Acesso amplo à rede
- Pool de recursos
- Elasticidade rápida
- Serviço mensurado

Temos notado constantemente cinco características ligadas a um aumento na performance organizacional e no desempenho de entrega de software. Este ano, analisamos se o uso da computação em nuvem era um preditor de uma infraestrutura mais flexível.

Infraestruturas flexíveis geram
30%
de aumento na
performance organizacional
em comparação a
infraestruturas inflexíveis

¹NIST Special Publication 800-145: "The NIST Definition of Cloud Computing".

	Porcentagem
Multicloud	19,6%
Nuvem pública	51,5%
Nuvem híbrida	33,6%
No local	19,8%
Sob o espaço de trabalho	3,2%
Outros	2,5%

* Os participantes podiam selecionar várias respostas.

Resultados

Mais uma vez, confirmamos algo que já havíamos descoberto: a maneira como a equipe usa a nuvem é um preditor mais forte de desempenho do que o simples uso da nuvem. Mesmo que o uso da nuvem seja importante para viabilizar o aprimoramento, isso não gera benefícios automaticamente. Na verdade, há indícios fortes de que a nuvem pública causa uma diminuição no desempenho do software e das operações, exceto para as equipes que usam uma infraestrutura flexível. Essa descoberta reforça ainda mais a ideia de que a “migração lift-and-shift” (o ato de passar cargas de trabalho do data center para a nuvem) não é benéfica e pode ser prejudicial.

O uso da computação em nuvem está associada a uma diminuição substancial no esgotamento, além de aumentos significativos na produtividade e na satisfação no trabalho.

Ambientes de computação

A tabela acima mostra onde os participantes disseram que executam o aplicativo ou serviço principal.



Tipo de nuvem	Performance organizacional	Desempenho da equipe	Desempenho de entrega de software	Desempenho operacional		
Privada		Nenhum sinal de impacto		Nenhum sinal de impacto		Aumento substancial associado ao uso da computação em nuvem
Pública		Aumento substancial associado ao uso da computação em nuvem		Diminuição substancial associada ao uso da computação em nuvem		Diminuição substancial associada ao uso da computação em nuvem
Híbrida		Aumento muito substancial associado ao uso da computação em nuvem		Diminuição substancial associada ao uso da computação em nuvem		Nenhum sinal de impacto
Multi				Nenhum sinal de impacto		Diminuição substancial associada ao uso da computação em nuvem

Apenas “usar a nuvem” gera resultados mistos

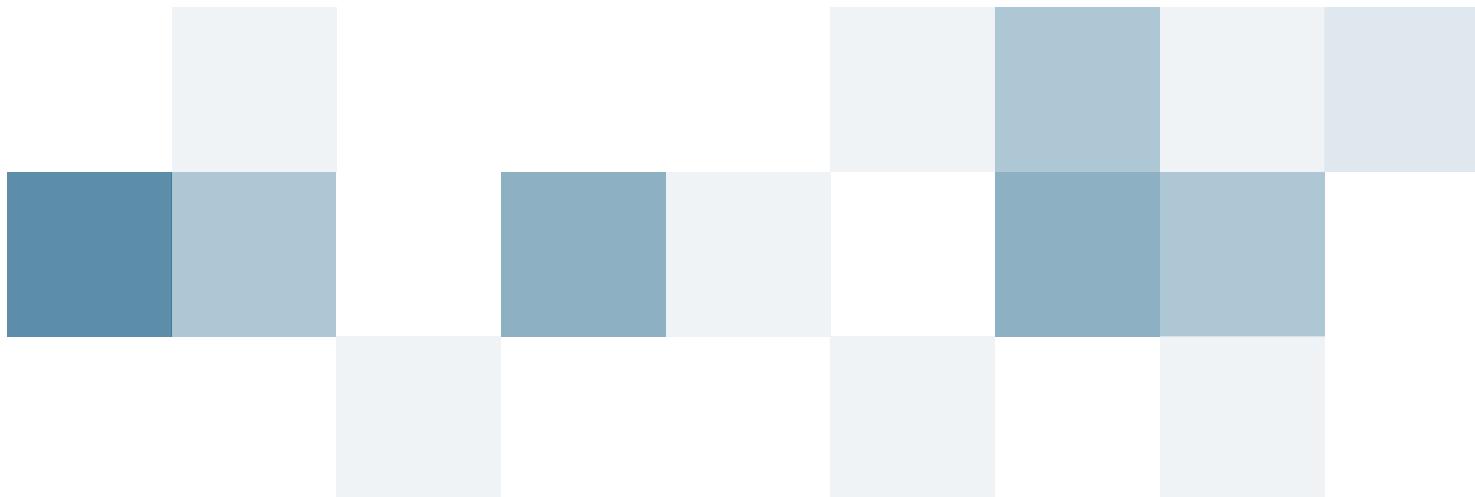
Conforme mostrado pelos resultados da tabela, apenas “usar a nuvem” tem impactos neutros ou negativos no desempenho operacional e de entrega de software.

Esse impacto entre neutro e negativo provavelmente se deve a profissionais que acabaram de começar a jornada em direção à nuvem e agora estão tendo que trabalhar em um ambiente novo, com outras ferramentas e fazendo algumas coisas de forma diferente. As empresas frequentemente usam a nuvem da mesma forma que utilizavam os data centers, mas com o aumento da complexidade e o peso cognitivo

de um ambiente novo. Caso a equipe não se adapte a esse novo ambiente, ela não vai conseguir melhorar o desempenho operacional nem o de entrega de software. Pelo contrário, esses desempenhos vão diminuir.

A única exceção a essa descoberta é o desempenho operacional no contexto da nuvem privada.

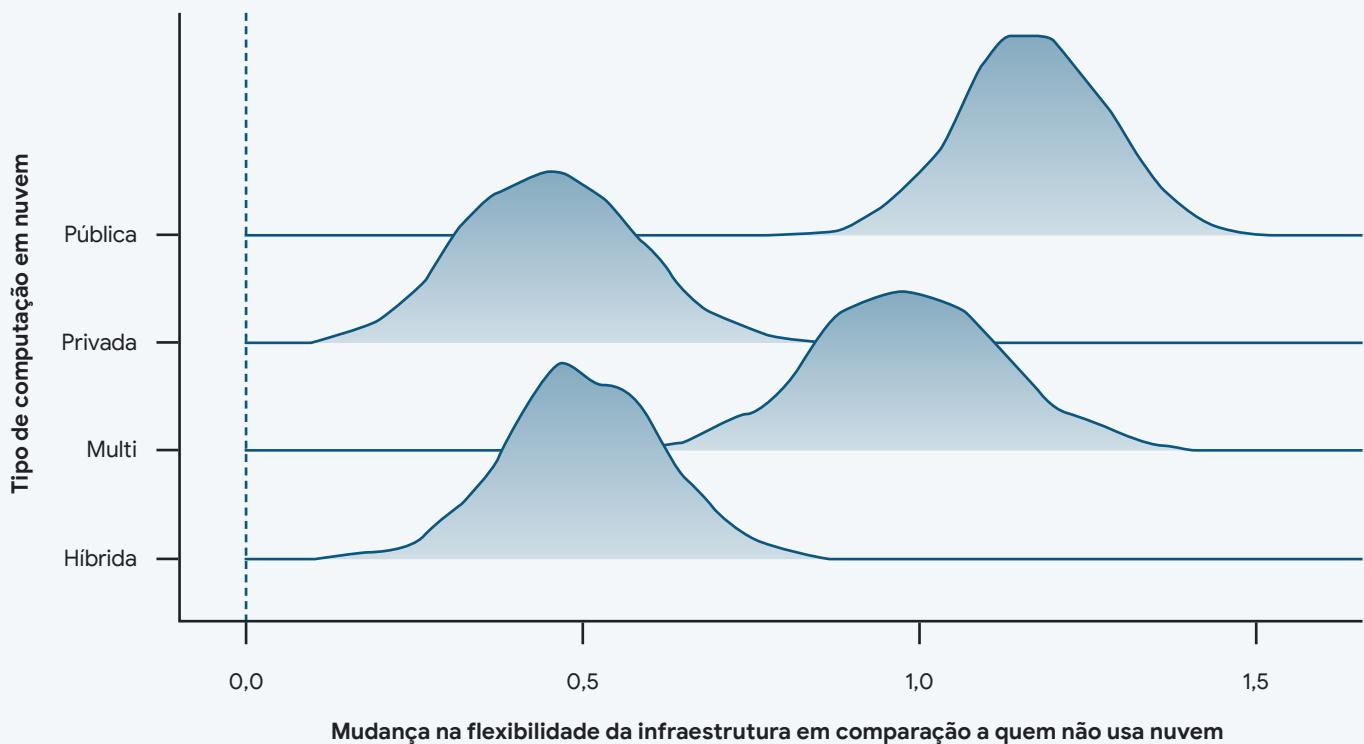
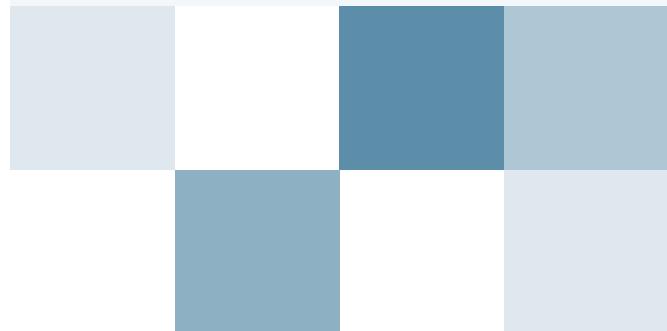
O que de fato melhora o desempenho operacional e de entrega de software é a **infraestrutura flexível**, que vamos discutir em breve.



A infraestrutura em nuvem viabiliza a flexibilidade

Usar uma nuvem pública torna a infraestrutura 22% mais flexível em comparação a infraestruturas sem nuvem. Usar várias nuvens também gera um aumento, porém menor do que o causado por uma única nuvem pública. E você já deve estar se perguntando: por quê? Nossos dados mostram que a infraestrutura flexível, muitas vezes viabilizada pela computação em nuvem, tem mais impacto do que o simples uso de uma plataforma de nuvem. Na maioria dos casos, a nuvem representa uma nova forma de fazer as coisas, e a perfeição leva tempo. Como cada plataforma de nuvem é diferente, quanto mais plataformas são usadas, mais esforço cognitivo é necessário para operá-las bem.

Usar uma nuvem pública torna a infraestrutura mais flexível em comparação a infraestruturas sem nuvem.



Infraestruturas flexíveis estão ligadas a um aumento no desempenho dos resultados principais

Capacidade	Performance organizacional	Desempenho da equipe	Desempenho de entrega de software	Desempenho operacional
Infraestrutura flexível	Aumento substancial associado a uma infraestrutura mais flexível	Aumento substancial associado a uma infraestrutura mais flexível	Aumento substancial associado a uma infraestrutura mais flexível	Aumento associado a uma infraestrutura mais flexível

É importante reconhecer que a infraestrutura flexível aumenta a performance organizacional e o desempenho da equipe, das operações e da entrega de software. Várias organizações optam por uma infraestrutura lift-and-shift para a nuvem, e isso pode ser uma ótima primeira etapa, mas é só o começo da jornada. Se você decidir adotar uma migração lift-and-shift para parte das suas cargas de trabalho, a próxima etapa será refatorá-las para modernizar e poder usar uma infraestrutura flexível.

A infraestrutura flexível que a computação em nuvem oferece tem um impacto positivo nos principais resultados

Tipo de nuvem com infraestrutura flexível	Performance organizacional	Desempenho da equipe	Desempenho de entrega de software	Desempenho operacional
Privada	Parcialmente mediado por uma infraestrutura flexível			
Pública	Totalmente mediado por uma infraestrutura flexível	Totalmente mediado por uma infraestrutura flexível	Parcialmente mediado por uma infraestrutura flexível	Parcialmente mediado por uma infraestrutura flexível
Híbrida	Parcialmente mediado por uma infraestrutura flexível	Totalmente mediado por uma infraestrutura flexível	Parcialmente mediado por uma infraestrutura flexível	Totalmente mediado por uma infraestrutura flexível
Multi	Parcialmente mediado por uma infraestrutura flexível			

Quando usadas para maximizar as características da infraestrutura flexível, as plataformas de computação em nuvem predizem um impacto positivo no desempenho operacional e de entrega de software. Essa diferença no impacto está relacionada ao que profissionais e líderes já sabem: apenas mudar as cargas de trabalho de um data center para a nuvem não é garantia de sucesso. A chave é aproveitar a infraestrutura flexível que a nuvem oferece.

Tipo de infraestrutura	Resultados
Nuvem com infraestrutura flexível	
Nuvem sem flexibilidade	

Para otimizar os benefícios, você precisa repensar a forma de criar, testar, implantar e monitorar seus aplicativos. Uma parte importante desse processo é aproveitar as cinco características da computação em nuvem: autoatendimento sob demanda, acesso amplo à rede, pool de recursos, elasticidade rápida e serviço mensurado.



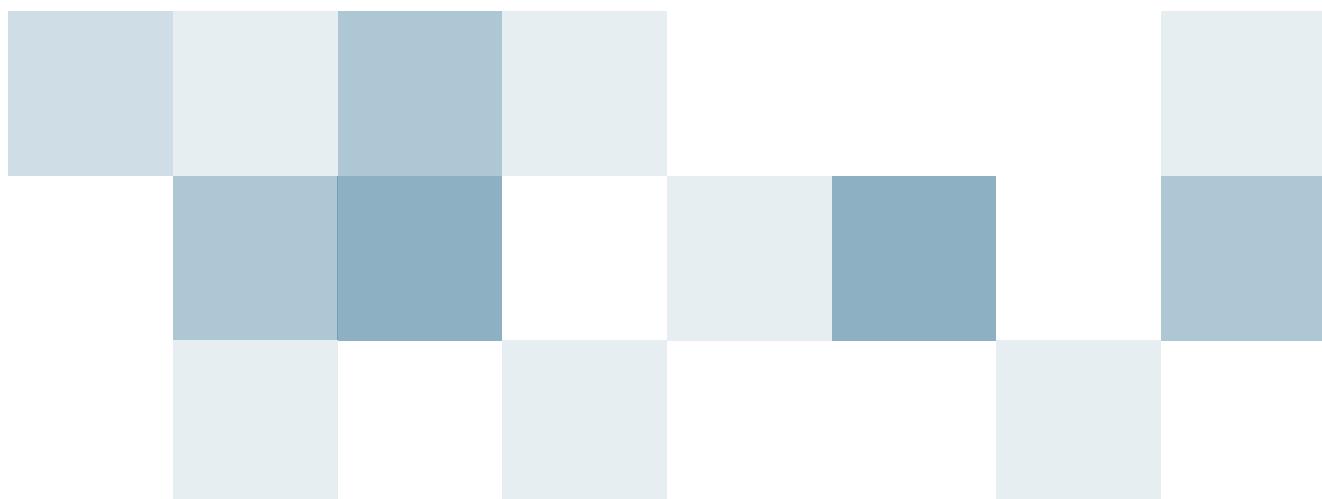
A computação em nuvem melhora o bem-estar

Tipo de nuvem	Esgotamento*	Satisfação no trabalho	Produtividade
Privada		Nenhum sinal de impacto	
Pública		Diminuição muito substancial associada ao uso da computação em nuvem	
Híbrida		Nenhum sinal de impacto	
Multi		Nenhum sinal de impacto	

*Note que o esquema de cores está trocado para o esgotamento. Isso foi feito porque diminuir o esgotamento é positivo!

Os dados mostram que a computação em nuvem é muito benéfica para o bem-estar dos funcionários. Constatamos um aumento substancial na produtividade e na satisfação no trabalho, além de um impacto neutro ou positivo no esgotamento. Em outras palavras, a nuvem não tem um efeito prejudicial no bem-estar, mesmo que a computação em nuvem exija mais esforço cognitivo, aprender a usar novas ferramentas e formas de trabalho.

Como profissionais, temos algumas hipóteses sobre por que isso acontece. Engenheiros gostam de aprender e resolver problemas, além de trabalhar em um ambiente com características de computação flexível. Aprender novas tecnologias não é só divertido, mas também uma ótima forma de avançar na carreira. Os engenheiros se sentem mais felizes quando a organização está tendo sucesso.



Nada disso funciona sem investir na cultura

Pontos principais

Cultura é um fator fundamental para elevar o bem-estar dos funcionários e a performance organizacional. Uma cultura saudável pode ajudar a reduzir o esgotamento, aumentar a produtividade e melhorar a satisfação no trabalho. Ela também causa um aumento significativo na performance organizacional e no desempenho operacional, da equipe e de entrega de software. Uma cultura organizacional saudável pode ajudar as equipes a implementar melhor as capacidades técnicas associadas a esses resultados.



Introdução

É difícil definir o que é cultura. Nós focamos em indicadores que podem nos informar algo sobre a experiência das pessoas no trabalho. Usamos a tipologia de Westrum para cultura organizacional¹ porque ela tem sido um preditor consistente de performance. E, neste ano, constatamos que organizações com uma cultura generativa (conforme a definição de Westrum) continuam tendo uma boa performance.

A tabela a seguir lista aspectos que nós acreditamos que contribuem para a cultura da organização e da equipe.

Aspecto	Definição
Cultura organizacional de Westrum	Como uma organização tende a reagir a problemas e oportunidades. Há três tipos de cultura: generativa, burocrática e patológica.
Estabilidade da organização	Quanto o ambiente é estável ou instável para os funcionários.
Estabilidade do emprego	Com que frequência os funcionários se preocupam com a estabilidade do emprego.
Flexibilidade	Como, onde e quando cada pessoa trabalha.
Compartilhamento de conhecimentos	Como ideias e informações se espalham por uma organização. Os membros da equipe respondem perguntas uma vez e a informação fica disponível para os outros. As pessoas não precisam esperar por respostas.
Centralidade do usuário	Desenvolver softwares com foco no usuário final e ter um conhecimento profundo das necessidades e metas dos usuários. Os indicadores dos usuários são utilizados para criar produtos e serviços melhores.
Distribuição do trabalho	Processos formais que ajudam as equipes a distribuir tarefas trabalhosas de maneira equitativa entre os membros.

A diferença entre aspectos culturais, capacidades processuais e capacidades técnicas nem sempre é clara. Acreditamos que a cultura surge das práticas e vice-versa. Vamos abordar isso mais tarde durante a discussão das descobertas.

¹<http://bmj.co/1BRGh5q>

O que descobrimos e o que isso significa?

Uma cultura saudável melhora resultados importantes

De modo geral, uma cultura saudável tem um impacto positivo em todos os resultados importantes. Replicamos as descobertas de anos anteriores que apontam que a cultura generativa melhora a performance organizacional e o desempenho operacional e de entrega de software. Ela também eleva a nova métrica deste ano: o desempenho da equipe.

Descobrimos que o desenvolvimento de software centrado no usuário causa um aumento significativo na performance. Isso vale a pena ser destacado. As organizações podem ter uma série de benefícios quando colocam o usuário em primeiro lugar. O feedback dos usuários ajuda as equipes a priorizar projetos e criar produtos e serviços que atendam às necessidades deles. Isso aprimora a experiência do usuário, melhora a satisfação do usuário e aumenta a receita.

Também medimos a distribuição do trabalho dentro das equipes para avaliar a integridade da cultura de uma organização. Descobrimos que uma distribuição equitativa do trabalho beneficia a performance organizacional e o desempenho da equipe. No entanto, descobrimos que a distribuição equitativa do trabalho está associada a um desempenho menor de entrega de software. Pode ser que os processos formais relacionados à distribuição de trabalhos façam com que tarefas complexas do pipeline de entrega de software levem mais tempo para serem concluídas. Também é possível que os processos formais afetem qual pessoa da equipe assume ou recebe uma determinada tarefa.

Outra descoberta aparentemente incongruente é que a estabilidade da organização causa uma diminuição

Equipes com cultura generativa têm

30%

mais performance organizacional em comparação com as demais

pequena (porém significativa) no desempenho de entrega de software. Uma possível explicação é que organizações mais consolidadas (e provavelmente maiores) não sentem a pressão de agir com a mesma velocidade que organizações mais novas, menos consolidadas (e menores). Pode ser que organizações mais consolidadas já tenham um produto estabelecido que permite que ela tenha uma velocidade de entrega de software mais flexível.

Quando as informações fluem com facilidade, a produtividade aumenta. Descobrimos que níveis altos de compartilhamento de informação estão associados a um aumento no desempenho operacional e de entrega de software. Quando as informações estão acessíveis e há poucos gargalos de conhecimento, as pessoas podem passar o tempo nas tarefas que importam em vez de buscar as informações que precisam para fazer essas tarefas.

Por fim, formatos de trabalho flexíveis, em que os funcionários podem determinar quando, onde e como trabalham, têm um impacto positivo em todas as métricas de desempenho. Isso é especialmente verdadeiro para o desempenho de entrega de software. Mesmo que as organizações implementem políticas de trabalho remoto mais rígidas, é possível que permitir um pouco de flexibilidade aos funcionários seja benéfico.

Aspecto cultural	Efeito no desempenho da equipe	Efeito na performance organizacional	Efeito no desempenho de entrega de software	Efeito no desempenho operacional
Cultura organizacional de Westrum		Aumento substancial		Aumento substancial
Estabilidade da organização		Aumento pequeno		Aumento substancial
Estabilidade do emprego		Aumento pequeno		Nenhum efeito
Flexibilidade		Aumento pequeno		Aumento substancial
Compartilhamento de conhecimentos		Aumento pequeno		Aumento substancial
Centralidade do usuário		Aumento substancial		Aumento substancial
Distribuição do trabalho		Aumento substancial		Nenhum efeito



Uma cultura saudável melhora as capacidades técnicas

Nossas descobertas sugerem que uma boa cultura ajuda a melhorar a implementação de capacidades técnicas. Acreditamos que a relação entre cultura e capacidades técnicas é recíproca: a cultura surge das práticas e vice-versa.

A cultura é ampla e difícil de definir, e as capacidades técnicas costumam ter um escopo específico e serem bem-definidas. Isso afeta como as pessoas da organização podem ajudar a impulsionar mudanças.

Por exemplo, os líderes podem criar estruturas de incentivo que promovem uma cultura generativa.

Líderes e colaboradores individuais podem dar ênfase a uma abordagem centrada no usuário para o desempenho de software. Colaboradores individuais

podem ajudar a impulsionar a implementação de capacidades técnicas que melhoraram o desempenho: desenvolvimento baseado em linha principal, integração contínua, práticas de confiabilidade e arquitetura com acoplamento flexível. Implementar essas capacidades técnicas não é fácil, e para que isso seja feito com êxito as pessoas precisam ter uma mente aberta e saber depender e aprender umas com as outras. Todos esses componentes fazem parte de uma cultura saudável. Equipes assim podem servir de exemplo umas para as outras dentro da organização, que por sua vez pode se sentir empoderada para motivar mudanças que usem as ferramentas que ela tem à disposição.

Mudanças duradouras e significativas surgem na cultura da organização com esforços simultâneos de cima para baixo e de baixo para cima voltados a gerar mudanças.

Aspecto cultural	Efeito no desenvolvimento baseado em linha principal	Efeito em práticas de confiabilidade	Efeito na integração contínua	Efeito na entrega contínua	Efeito na arquitetura com acoplamento flexível
Cultura organizacional de Westrum	Aumento substancial	Aumento substancial	Aumento substancial	Aumento substancial	Aumento substancial
Estabilidade da organização	Aumento pequeno	Aumento substancial	Nenhum efeito	Nenhum efeito	Nenhum efeito
Estabilidade do emprego	Diminuição pequena	Diminuição pequena	Nenhum efeito	Nenhum efeito	Nenhum efeito
Flexibilidade	Nenhum efeito	Diminuição pequena	Aumento substancial	Aumento pequeno	Aumento substancial
Compartilhamento de conhecimentos	Nenhum efeito	Nenhum efeito	Nenhum efeito	Aumento pequeno	Aumento pequeno
Centralidade do usuário	Aumento substancial	Aumento substancial	Aumento substancial	Aumento substancial	Aumento substancial
Distribuição do trabalho	Aumento substancial	Aumento substancial	Aumento substancial	Aumento substancial	Aumento substancial

Uma cultura saudável melhora o bem-estar dos funcionários

Uma cultura saudável leva a altos níveis de bem-estar entre os funcionários porque reduz o esgotamento, aumenta a satisfação no trabalho e aumenta a produtividade. O bem-estar dos funcionários não é só um bônus agradável, mas sim um elemento fundamental da integridade e do sucesso geral da organização.

O que acontece quando as organizações não investem em uma cultura melhor? A probabilidade é que o esgotamento aumente e a satisfação no trabalho diminua. Os funcionários ficam cínicos e a produtividade cai. A saúde física e psicológica deles também é prejudicada.^{2,3} O esgotamento é persistente, não algo que as pessoas superam depois de tirar uma folga. O esgotamento aumenta a rotatividade porque faz os funcionários saírem para buscar ambientes de trabalho mais saudáveis.⁴ Por isso, é necessário realizar mudanças organizacionais que atenuem as causas do esgotamento.

Aspecto cultural	Efeito no esgotamento*	Efeito na satisfação no trabalho	Efeito na produtividade
Cultura organizacional de Westrum			
Estabilidade da organização			
Estabilidade do emprego			
Flexibilidade			
Compartilhamento de conhecimentos			
Centralidade do usuário			
Distribuição do trabalho			

*Note que o esquema de cores está trocado para o esgotamento. Isso foi feito porque diminuir o esgotamento é positivo!

² Adam Bayes, Gabriela Tavella & Gordon Parker (2021) "The biology of burnout: Causes and consequences", The World Journal of Biological Psychiatry, 22:9, 686–698. DOI: 10.1080/15622975.2021.1907713. <https://doi.org/10.1080/15622975.2021.1907713>

³ Maslach C, Leiter MP. "Understanding the burnout experience: recent research and its implications for psychiatry". World Psychiatry, junho de 2016, 15(2), 103–11. DOI: 10.1002/wps.20311. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4911781/>. PMID: 27265691; PMCID: PMC4911781.

⁴ L.A. Kelly, et al. "Impact of nurse burnout on organizational and position turnover". Nursing Outlook, janeiro de 2021, 96–102, janeiro/fevereiro de 2021, 96–102. DOI: doi.org/10.1016/j.outlook.2020.06.008

Capítulo 8

A importância de "como", "quando", "por quê" e "quem"

Pontos principais

Quem você é importa: descobrimos que alguns grupos de participantes têm resultados diferentes dos demais, como mais esgotamento ou menos produtividade.

Também identificamos práticas específicas para mitigar alguns desses resultados negativos.

Introdução

A análise de 2022 constatou um fenômeno: o formato do trabalho pode afetar o bem-estar de algumas pessoas, mas não todas.

Em 2022, descobrimos que as pessoas que se identificam como sub-representadas relataram ter níveis mais altos de esgotamento.¹ Neste capítulo vamos analisar como essa descoberta se repete e começar a entender por que grupos sub-representados são mais propensos a sentir esgotamento e quais fatores podem ajudar a evitar isso.

Além disso, a instabilidade que assolou diversos setores levaram muitos a questionarem novas contratações. As organizações estão preocupadas porque novos funcionários demoram muito tempo para se tornarem produtivos. Elas estão buscando formas de ajudar esses funcionários a chegarem ao mesmo nível que os outros mais rapidamente. Também vamos investigar isso aqui.

O que descobrimos e o que isso significa?

Algumas pessoas estão mais esgotadas que outras

No ano passado, descobrimos que mulheres ou pessoas de gênero autodeterminado ou que se consideravam sub-representadas relataram níveis mais altos de esgotamento do que homens ou pessoas que não se consideravam sub-representadas. Essas descobertas são consistentes com pesquisas anteriores que sugerem que pessoas sub-representadas têm mais esgotamento² e estresse relacionado ao trabalho³ do que colegas com mais representação.

Por isso, queríamos saber se essa diferença de esgotamento seria constatada de novo nos dados deste ano, e foi. Participantes que se identificaram como mulheres ou pessoas com gênero autodeterminado relataram ter níveis de esgotamento 6% mais altos do que os participantes homens. Participantes que se identificaram como sub-representados de alguma forma relataram níveis de esgotamento 24% mais altos do que participantes que não se identificaram como sub-representados.

¹ 2022 Accelerate State of DevOps Report. <https://dora.dev/research/2022/dora-report/>

² Sigalit Ronen and Ayala Malach Pines, "Gender Differences in Engineers' Burnout", Equal Opportunities International, 7 de novembro de 2008, <https://www.emerald.com/insight/content/doi/10.1108/02610150810916749/full/html>

³ Dalessandro C; Lovell A; Tanner OC; "Race, Marginalization, and Perceptions of Stress Among Workers Worldwide Post-2020". Sociological Inquiry, 3 de agosto de 2023, <https://onlinelibrary.wiley.com/doi/10.1111/soin.12505>

Alguns tipos de trabalho estão mais associados a esgotamento

Descobrimos que aspectos do espaço de trabalho que podem parecer neutros ou benéficos, como uma documentação de qualidade ou uma equipe estável, não reduzem o esgotamento de todas as pessoas (consulte [Capítulo 4: O papel fundamental da documentação](#)).

É possível que isso ocorra porque essas tarefas beneficiam a organização, mas contribuem para o esgotamento de algumas pessoas.⁴

Para entender a experiência de esgotamento dos participantes, perguntamos qual trabalho eles fazem. Isso foi medido de duas maneiras, analisando:

- Tarefas específicas, como programar, fazer reuniões e apoiar colegas.
- Características do trabalho, como falta de planejamento, visibilidade ou quantidade de tarefas repetitivas.

As características do trabalho são importantes porque a mesma tarefa pode ser encarada de modos distintos de acordo com a pessoa ou a época. Por exemplo, algumas revisões de código podem incluir tarefas repetitivas não planejadas, e outras podem ter alta visibilidade dentro da equipe porque demonstram liderança e experiência técnica.⁵

Participantes que se identificam como sub-representados relatam que fazem 24% mais tarefas repetitivas do que os não sub-representados. Participantes que se identificaram como mulheres ou pessoas com gênero autodeterminado relataram que fazem 40% mais tarefas repetitivas do que participantes que se identificam como homens. Esses dois grupos de participantes também relatam que o trabalho é menos planejado, não tem tanta visibilidade quanto o dos colegas e não condiz diretamente ao conjunto de habilidades profissionais deles. Essas descobertas explicam parcialmente o esgotamento que esses grupos relatam.

⁴ Linda Babcock, Brenda Peyser, Lise Vesterlund e Laurie Weingart. *The No Club* (Nova York: Simon and Schuster, 2022), 17.

⁵ Murphy-Hill, E. et al. "Systemic Gender Inequities in Who Reviews Code", Computer Supported Cooperative Work (2023) (no prelo), <https://research.google/pubs/pub52204>

⁶ Babcock et al., *The No Club*, 17.

Tarefas que não geram promoções

No livro *The No Club*, Babcock et al. descrevem um tipo específico de trabalho: "Uma tarefa que não gera promoção é importante para a organização, mas não vai ajudar você a progredir na carreira".⁶

Aqui, a palavra *progredir* é usada de modo abrangente: pode ser um aumento na remuneração ou mais experiência desejável para outros empregos.

Evidências mostram que mulheres realizam mais esse tipo de trabalho. Babcock et al. descrevem por que existe essa distribuição desigual do trabalho. Por exemplo, é mais comum que mulheres recebam esse tipo de trabalho e aceitem, porque existe um custo social caso neguem.

Elas também descrevem as consequências da distribuição desigual de tarefas. Por exemplo, algumas mulheres:

- Notam impactos negativos na própria carreira ou remuneração.
- Cumprem mais horas para ter um volume de trabalho relevante para a carreira delas.

Processos formais de distribuição de trabalho reduzem o esgotamento para alguns participantes

Perguntamos aos participantes se havia processos formais para distribuir o trabalho de forma justa.

Chamamos isso de **distribuição do trabalho**, e a hipótese era que isso mitigaria o esgotamento de alguns participantes.

Descobrimos que a distribuição de trabalho de fato reduz o esgotamento para participantes que se identificam como homens, mulheres ou pessoas com gênero autodeterminado. Em casos com alto nível de distribuição de trabalho, a diferença de esgotamento entre gêneros desaparecia.

Tivemos a surpresa de descobrir que a distribuição do trabalho não tinha impacto no esgotamento dos participantes que se identificaram como sub-representados. Essa descoberta gera mais dúvidas: será que processos formais para distribuir o trabalho de forma justa ainda resultam em uma distribuição desigual? Será que o “trabalho igualitário” considera as características das tarefas, como interrupções ou visibilidade? E como podemos mitigar outros fatores que contribuem para o esgotamento, para além das tarefas, que possam ser mais significativos para esse grupo?



Como fomentar uma cultura de pertencimento

Autor: Dr. Jeffrey Winer, psicólogo e professor assistente do Boston Children's Hospital da Harvard Medical School

Uma descoberta importante e um pouco de contexto

Uma descoberta importante deste relatório é que pessoas que se definem como sub-representadas têm muito mais esgotamento do que os colegas. O relatório investigou algumas possíveis causas disso. Nesta seção, queremos conectar essas descobertas a uma pesquisa mais ampla sobre pertencimento e estratégias associadas de prática organizacional.

Identificar-se como sub-representado em um grupo demonstra uma vulnerabilidade à “incerteza de pertencimento”,⁷ um fenômeno psicológico bem estabelecido (consulte [“Understanding and Overcoming Belonging Uncertainty” do Dr. Geoffrey Cohen](#)).⁸ Essa incerteza (por exemplo, “Será que eu pertenço a este lugar?”, “Será que pessoas como eu podem ter sucesso aqui?”) é reforçada ou redefinida pelas experiências contínuas das pessoas e como elas interpretam essas experiências. Esses processos consolidados relacionados à incerteza de pertencimento podem ajudar a contextualizar o que descobrimos neste relatório: pessoas que se identificam como sub-representadas relatam ter níveis mais altos de esgotamento.

O que as organizações podem fazer?

É importante lembrar que diversidade, inclusão, equidade e pertencimento são coisas diferentes e exigem estratégias distintas, interconectadas e repetidas para serem alcançadas. É necessário fazer compromissos verdadeiros e repetidos para garantir o pertencimento.

Se as pessoas estiverem tendo problemas na organização, a primeira pergunta a ser feita não é “Qual é o problema da pessoa?”, mas sim “Por que uma pessoa se sentiria assim e quais elementos estruturais da nossa organização facilitam esse sentimento (por exemplo, quais elementos fazem esse sentimento perdurar ou piorar)?”.

Depois de identificar os problemas, é necessário implementar mudanças na organização e oferecer suporte individual, ou seja, adotar uma abordagem com duas frentes. Oferecer apoio para que as pessoas mudem os sistemas que regem uma organização é uma forma de fazer essas mudanças serem incorporadas ao sistema e durarem mais tempo que agentes individuais. Essa mentalidade sistemática e sustentável permite que as mudanças sejam incorporadas pela instituição e durem mais tempo que agentes individuais. É graças a essa qualidade generativa que as organizações podem fomentar o pertencimento. A palavra *fomentar* é importante aqui. O pertencimento é fruto de experiências e ações repetidas, mas esse esforço nunca acaba. Por isso, ele é fundamental para a integridade e a produtividade do espaço de trabalho.

Existem várias ferramentas disponíveis para ajudar as organizações nesse sentido. Por exemplo, o relatório da Surgeon General de 2023 sobre solidão identificou que o vínculo social e o pertencimento são antídotos essenciais para a solidão e o esgotamento.⁹

⁷ Walton GM; Cohen GL; “A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students”, Science (Nova York, NY), acessado em 20 de setembro de 2023, <https://pubmed.ncbi.nlm.nih.gov/21415354/>

⁸ <https://behavioralscientist.org/understanding-and-overcoming-belonging-uncertainty/>

⁹ Office of the Assistant Secretary for Health (OASH), “New Surgeon General Advisory Raises Alarm about the Devastating Impact of the Epidemic of Loneliness and Isolation in the United States”, HHS.gov, 3 de maio de 2023, <https://www.hhs.gov/about/news/2023/05/03/new-surgeon-general-advisory-raises-alarm-about-devastating-impact-epidemic-loneliness-isolation-united-states.html>

O desafio da produtividade para novos contratados

Novos contratados (com menos de um ano de experiência na equipe) têm uma produtividade 8% mais baixa do que colegas com mais de um ano de experiência. Isso já era esperado. Entrar em uma nova equipe é desafiador e, mesmo que tenha experiência no cargo, a quantidade de conhecimento específico sobre a equipe que você precisa para entrar no ritmo pode ser intimidadora. Além disso, participar de uma equipe não é só uma questão de habilidade e conhecimento. Também há um componente social que é fundamental para a produtividade. Aspectos como o pertencimento, a sensação de estar contribuindo e a segurança psicológica levam tempo para serem desenvolvidos.

Existe algo para ajudar novos contratados a acelerarem esse processo?

Nossa hipótese é que as organizações têm três formas de ajudar novos contratados:

- Disponibilizar documentos de alta qualidade.
- Incorporar a inteligência artificial a fluxos de trabalho, o que demonstradamente ajuda mais funcionários inexperientes do que experientes.
- Incentivar o trabalho presencial colaborativo, o que alguns sugerem ser especialmente benéfico na etapa de integração.



Os resultados deste ano sugerem que uma documentação de alta qualidade causa impactos significativos na produtividade (confira o [Capítulo 4: O papel fundamental da documentação](#)) e a IA traz pequenos benefícios à produtividade individual (confira o [Capítulo 3: Capacidades técnicas que predizem o desempenho](#)). Não temos motivo para esperar que esses efeitos não se estendam a novos contratados. Quando analisamos os dados, é isso que constatamos: os novos contratados se beneficiam dessas práticas, mas não mais ou menos que as outras pessoas. Em outras palavras, essas práticas não dão apoio especial para novos contratados.

Se quiser ajudar novos contratados e todas as outras pessoas, a documentação de alta qualidade é um ótimo começo por causa da clareza e sustentabilidade do efeito que ela tem na produtividade. Vale a pena comentar que novos contratados de equipes com uma documentação bem redigida (1 desvio padrão acima da média) são 130% mais produtivos que os de equipes com documentação má redigida (1 desvio padrão abaixo da média).

A discussão sobre retorno ao presencial será breve para evitar mais polêmicas. Discutimos a importância da flexibilidade no [Capítulo 7: Nada disso funciona sem investir na cultura](#). Além disso, nossos dados não são experimentais e, mesmo tentando evitar vieses, os benefícios do formato de trabalho são uma questão complexa e sociologicamente rica, então é difícil chegar a conclusões definitivas (é bom ter isso em mente ao ler pesquisas ou artigos sobre esse tema). O que os dados mostram claramente é que a flexibilidade tem um impacto positivo na produtividade. O que não está claro é se o local onde o trabalho é realizado tem o mesmo impacto.

Isso também é verdadeiro para novos contratados. Não encontramos evidência de que o trabalho presencial tenha um benefício específico para novos contratados. Se você quiser otimizar a produtividade, oferecer aos novos contratados a flexibilidade de definir como, onde e quando trabalhar parece ser uma aposta mais garantida do que forçá-los a estar no escritório. É claro que as organizações não são (e provavelmente não deveriam ser) otimizadas pensando apenas na produtividade. Também pensamos na produtividade em termos de trabalho que agrega valor, não apenas produtos (linhas de código), e trabalho que não gera esgotamento ou tarefas repetitivas.



Posfácio

Considerações finais

Agradecemos por participar da pesquisa deste ano e por ler este relatório. Estamos sempre em busca das melhores formas de investigar conexões entre o trabalho das equipes e os resultados que elas conseguem atingir.

A lição mais importante depois de tantos anos neste programa de pesquisa é que as equipes que adotam uma mentalidade e práticas de aprimoramento contínuo conseguem atingir os melhores resultados.

As capacidades que analisamos podem ser usadas como comandos para potencializar os resultados. Alguns desses comandos estão ao alcance das pessoas, enquanto outros só podem ser usados com o esforço coordenado de toda a organização. Identifique quais comandos precisam ser ajustados na organização e depois invista nesses ajustes.

O trabalho de aprimoramento não acaba, mas pode ser decisivo para o sucesso a longo prazo de pessoas, equipes e organizações. A responsabilidade de incentivar o trabalho de aprimoramento é compartilhada entre líderes e profissionais.

Como você vai colocar a pesquisa em prática?

Investigue estas descobertas no contexto da sua organização, equipes e serviços que vocês prestam para os clientes.

Participe da comunidade DORA para compartilhar suas experiências, aprender com outras pessoas e se inspirar em quem também está em uma jornada de aprimoramento contínuo: <https://dora.community>.

Agradecimentos

Todo ano, este relatório recebe o suporte de uma família grande de colaboradores apaixonados do mundo todo. Todas as etapas envolvidas na produção, como elaboração das perguntas da pesquisa, localização, análise, escrita, edição e formatação, são resultado do enorme esforço de colegas que nos ajudaram a tornar isso uma realidade. Os autores gostariam de agradecer a todas essas pessoas pelas opiniões, orientações e companheirismo.

Colaboradores

Equipe principal

James Brookbank

Kim Castillo

Derek DeBellis

Nathen Harvey

Michelle Irvine

Amanda Lewis

Eric Maxwell

Steve McGhee

Dave Stanke

Kevin Storer

Daniella Villalba

Brenna Washington

Editores

Mandy Grover

Jay Hauser

Stan McKenzie

Anna Eames Mikkawi

Mike Pope

Tabitha Smith

Olinda Turner

Localização da pesquisa

Daniel Amadei

Kuma Arakawa

William Bartlett

Antonio Guzmán

Shogo Hamada

Yuki Iwanari

Vincent Jobard

Gustavo Lapa

Mauricio Meléndez

Jeremie Patonnier

Miguel Reyes

Pedro Sousa

Laurent Tardif

Kimmy Wu

Vinicius Xavier

Yoshi Yamaguchi

Consultores e especialistas da área

Jared Bhatti

Lisa Crispin

Rob Edwards

Dave Farley

Steve Fenton

Dra. Nicole Forsgren

Aaron Gillies

Denali Lumma

Emerson Murphy-Hill

Harini Sampath

Robin Savinar

Dustin Smith

Jess Tsimeris

Dra. Laurie Weingart

Betsalel (Saul) Williamson

Dr. Jeffrey Winer

Patrocinadores



Autores



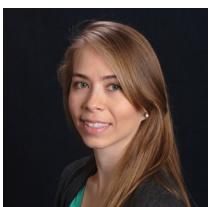
Derek DeBellis

Derek é pesquisador quantitativo de experiência do usuário no Google e pesquisador principal da DORA. Derek foca em fazer pesquisas, analisar registros e descobrir maneiras de medir conceitos que demonstram que um produto ou recurso agrega valor de verdade para as pessoas. Derek tem publicações sobre a interação entre humanos e IA, o impacto do surgimento da COVID-19 sobre a queda no tabagismo, o design para erros de PLN, o papel da UX nas discussões sobre privacidade, a cultura de equipes e a relação entre IA e o bem-estar e a produtividade dos funcionários. Sua pesquisa extracurricular atual é investigar formas de simular a propagação de crenças e poder.



Amanda Lewis

Amanda Lewis é líder de desenvolvimento da DORA.community e engenheira de relações entre desenvolvedores na equipe DORA Advocacy no Google Cloud. Ela passou a carreira criando vínculos entre operadores e gerentes de produto e desenvolvendo a gestão de projetos e a liderança. Ela trabalhou em equipes de desenvolvimento de plataformas de e-commerce, sistemas de gerenciamento de conteúdo, ferramentas de observabilidade e apoio a desenvolvedores. Esses vínculos e conversas ajudaram a deixar os clientes mais felizes e conquistar resultados melhores para os negócios. Ela traz experiência e simpatia que ajudam as equipes a entender e implementar práticas de confiabilidade e entrega de software.



Daniella Villalba

Daniella Villalba é pesquisadora de experiência do usuário no Google. As pesquisas dela buscam entender quais fatores afetam a felicidade e a produtividade de desenvolvedores. Antes do Google, Daniella estudou os benefícios da meditação e os fatores psicossociais que afetam as experiências de estudantes universitários. Ela é doutora em psicologia experimental pela Universidade Internacional da Flórida.



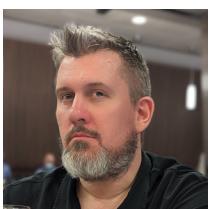
Dave Farley

Dave Farley é diretor-gerente e fundador da Continuous Delivery Ltd, autor de *Modern Software Engineering* e coautor do best-seller *Continuous Delivery*. Ele é um dos autores do *Reactive Manifesto* e vencedor do Prêmio Duke pelo projeto de código aberto LMAX Disruptor. Dave é pioneiro em entrega contínua, formador de opinião e especialista em CD, DevOps, desenvolvimento conduzido por testes (TDD, na sigla em inglês) e design de software. Também tem longo histórico de criação de equipes com alto desempenho, preparação de organizações para o sucesso e criação de softwares excepcionais. Dave tem o compromisso de compartilhar sua experiência e técnica com desenvolvedores do mundo todo para ajudar a melhorar o design, a qualidade e a confiabilidade dos softwares. Ele compartilha isso em consultorias,¹, no canal dele no YouTube² e em treinamentos.



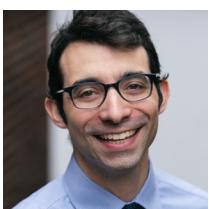
Eric Maxwell

Eric Maxwell lidera a área de transformação digital de DevOps do Google, em que orienta as melhores empresas do mundo a entregar valor mais rápido. Eric passou a primeira metade da carreira como engenheiro e enfrentou muitos desafios, automatizando tudo e adquirindo empatia por outros profissionais. Ele é cocriador do Programa de Modernização de Aplicativos Google Cloud (CAMP, na sigla em inglês) e membro da equipe da DORA. Antes do Google, Eric trabalhava com outros fãs de trocadilhos na Chef Software.



James Brookbank

James Brookbank é arquiteto de soluções em nuvem no Google. Arquitetos de soluções ajudam os clientes do Google Cloud a resolver problemas técnicos complexos e oferecer orientações especializadas sobre arquitetura. Antes de se juntar ao Google, James trabalhou em várias empresas de grande porte com foco na infraestrutura de TI e em serviços financeiros.



Dr. Jeffrey Winer

Jeffrey P. Winer, Ph.D., é psicólogo, consultor de sistemas de saúde comportamental, desenvolvedor de tratamentos psicossociais no Centro de Resiliência da Comunidade e Trauma (TCRC, na sigla em inglês) do Boston Children's Hospital e professor assistente da Faculdade de Medicina de Harvard. Ao lado dos colegas do TCRC, o trabalho dele foca em criação, testes, disseminação e implementação de intervenções psicossociais informadas pelo trauma e culturalmente responsivas para jovens e famílias de imigrantes e refugiados. Ele é coautor do livro *Mental Health Practice with Immigrant and Refugee Youth: A Socioecological Framework*.³ Ele já fez consultorias para programas dos Estados Unidos e Canadá. As ferramentas de prevenção e intervenção psicossocial que ele ajudou a desenvolver ou adaptar são usadas no mundo todo. Para saber mais, acesse <http://www.drjeffwiner.com>.

¹ <https://continuous-delivery.co.uk/engineering-for-software>

² <https://www.youtube.com/@ContinuousDelivery>

³ <https://www.apa.org/pubs/books/4317536>



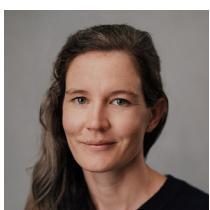
Kevin Storer

Kevin M. Storer é pesquisador de experiência do usuário no Google, onde lidera pesquisas que buscam entender como equipes de desenvolvimento de software interagem com ferramentas de DevOps e com o uso delas. Antes de se juntar ao Google, Kevin fez um Doutorado em Informática na Universidade da Califórnia em Irvine. Nos setores público e privado, Kevin já publicou trabalhos de alto impacto sobre programação centrada em humanos, experiência do desenvolvedor, comportamento informacional, acessibilidade e computação ubíqua.



Kim Castillo

Kim Castillo é gerente do programa de experiência do usuário no Google. Kim lidera as iniciativas interfuncionais na DORA, desde a supervisão de operações de pesquisa até a publicação deste relatório. Kim também trabalha com pesquisa de UX para Duet AI no Google Cloud. Antes do Google, Kim tinha uma carreira em entrega de software e atuava com gerenciamento de programas técnicos e coaching ágil. As bases de Kim são pesquisas psicosociais que focam nos temas de assassinatos extrajudiciais, aumento da pobreza urbana, trauma comunitário e resiliência em seu país natal, as Filipinas. A DORA combina as diversas paixões de Kim: práticas de entrega de software, segurança psicológica e organizações descentralizadas.



Michelle Irvine

Michelle Irvine é uma escritora de conteúdo técnico do Google que lidera pesquisas sobre a produção e o impacto da documentação técnica. Antes do Google, ela atuou no setor editorial da área de educação e como escritora de conteúdo técnico para um software de simulação física. Michelle é formada em Física e mestre em Retórica e Design de Comunicação na Universidade de Waterloo.



Nathen Harvey

Nathen Harvey é líder da equipe DORA Advocacy como gerente de engenharia de relacionamento com desenvolvedores no Google Cloud. Nathen teve o privilégio de trabalhar com algumas das melhores equipes e comunidades de código aberto, ajudando a aplicar os princípios e as práticas de DevOps e SRE. Ele é coautor do relatório *Accelerate State of DevOps* há três anos. Nathen também coeditou e contribuiu para o livro *97 Things Every Cloud Engineer Should Know*.



Steve McGhee

Steve McGhee é um defensor da confiabilidade que ajuda equipes a entenderem a melhor forma de criar e operar serviços confiáveis de alto nível. Antes disso, ele passou mais de 10 anos como engenheiro de confiabilidade do site no Google aprendendo a escalar sistemas globais na Pesquisa Google, YouTube, Android e Google Cloud. Ele gerenciou várias equipes de engenharia na Califórnia, no Japão e no Reino Unido. Steve também passou algum tempo em uma empresa californiana para ajudá-la a passar para a nuvem.

Metodologia

Este capítulo descreve como este relatório começou como uma série de ideias iniciais e se tornou o que é hoje (com ideias um pouco mais completas). Esperamos que ele solucione várias dúvidas sobre como este relatório foi elaborado e ajude você a fazer suas próprias pesquisas.

Etapa 1. Elaborar um conjunto de resultados que consideramos importantes para organizações com alto desempenho e baseadas em tecnologia

Isso é fundamental. Nossa programa é baseado em um guia para ajudar as pessoas alcançarem o que consideram importante. Então, precisamos saber aonde as pessoas, organizações ou equipes querem chegar. Como descobrimos isso? Usamos uma mistura de pesquisa qualitativa (ou seja, perguntar o que pessoas, equipes e organizações querem conquistar), questionários, interações com a comunidade geral e vários workshops. Chegamos a alguns resultados consistentes, como:

- **Performance organizacional.** A organização não deve gerar somente receita, mas também valor para clientes e para a comunidade como um todo.
- **Desempenho da equipe.** A habilidade de uma equipe de serviço ou aplicativo gerar valor, inovar e colaborar.
- **Bem-estar dos funcionários.** As estratégias que uma organização ou equipe adotam devem beneficiar os funcionários, reduzindo o esgotamento, aumentando a satisfação no trabalho e melhorando a capacidade das pessoas de alcançar resultados valiosos (ou seja, ter produtividade).

Também ouvimos as pessoas falarem sobre metas como:

- **Desempenho de entrega de software.** As equipes conseguem implantar software de forma ágil e bem-sucedida.
- **Desempenho operacional.** O software enviado oferece uma experiência confiável para o usuário.



Etapa 2. Elaborar hipóteses sobre como, quando e por que esses resultados são alcançados

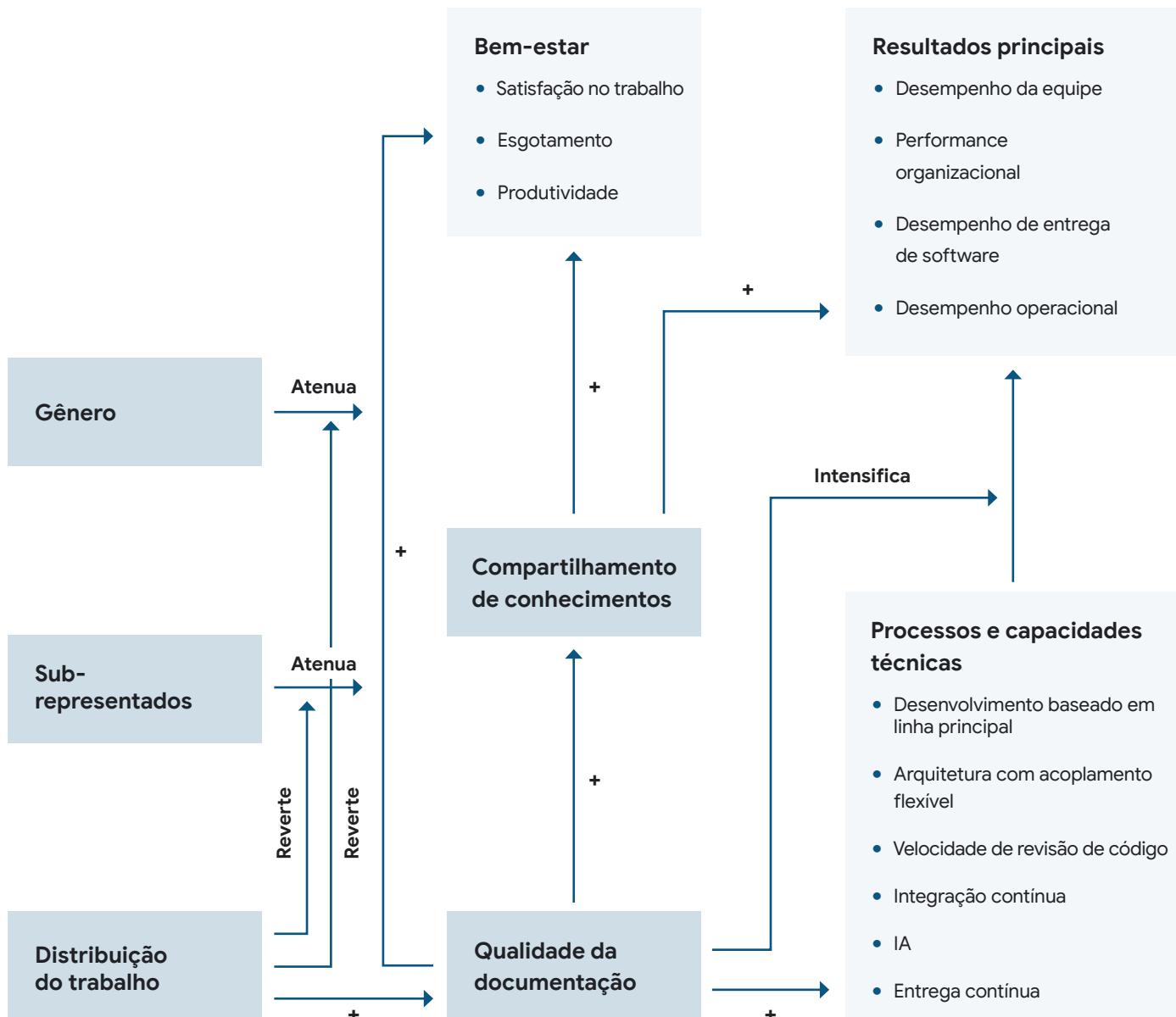
Pensando nos resultados definidos na etapa 1, precisamos elaborar hipóteses sobre quais caminhos as equipes podem seguir para chegar até eles. Isso envolve procurar fatores que parecem ter um impacto consistente em um resultado. Queremos afirmar que “Se tudo ficar igual, x vai afetar y”. Essa informação pode ajudar os profissionais a tomarem decisões informadas sobre que tipo de mudança testar.

Não queremos somente entender quais ações causam impacto; queremos investigar sob quais condições essas ações têm mais ou menos impacto. Isso significa perguntar “quando” e “para quem”. Está comprovado que, no geral, a documentação de qualidade reduz drasticamente o esgotamento. Mas, quando analisamos os participantes que relataram ser sub-representados, o oposto é verdadeiro: a documentação de qualidade *aumenta* o esgotamento. Entender as condições desses efeitos é fundamental porque é raro que uma equipe ou pessoa se encaixe no que é considerada a média geral.

Além disso, elaboramos hipóteses sobre quais mecanismos explicam o *porquê* ou *como*. Este ano, elaboramos hipóteses sobre por que pessoas que se identificam como sub-representadas sentem mais esgotamento, como mostrado pelos resultados do ano passado e por uma ampla literatura sobre o tema. A primeira pergunta que as pessoas fazem é “Por que isso está acontecendo?”. Para tentar responder essa pergunta, pensamos em quais mecanismos devem ser testados. Por exemplo, pessoas que se identificam como sub-representadas podem sentir mais esgotamento porque assumem (ou recebem) mais tarefas repetitivas.



Em seguida, essas hipóteses foram mapeadas para servirem de base para nossas pesquisas e análises. Aqui está um exemplo de um modelo hipotético do capítulo sobre documentação, com efeitos, condicionalidade e mecanismo:

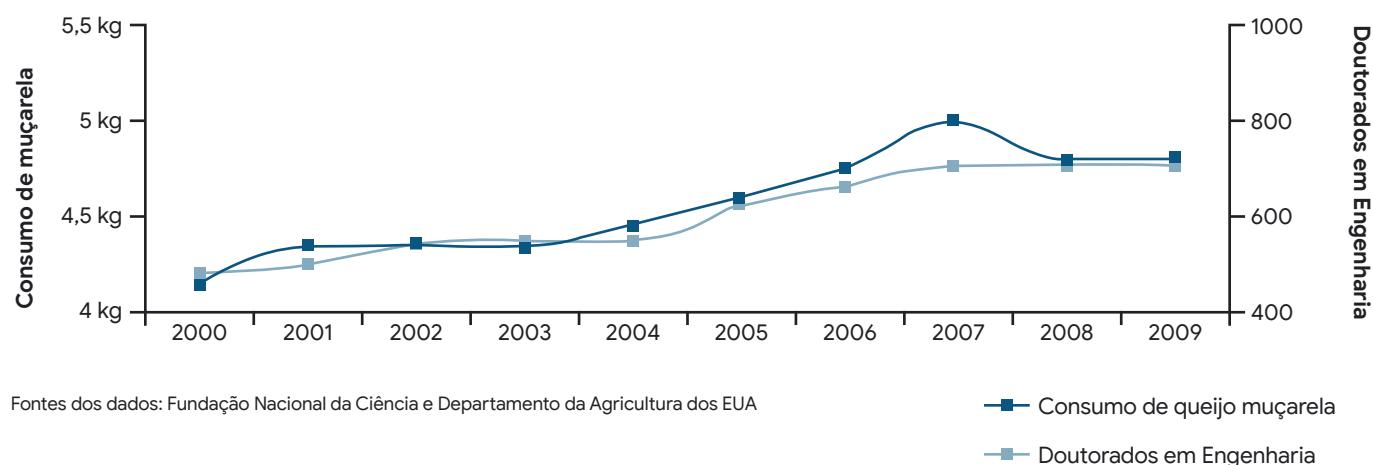


Etapa 3. Elaborar hipóteses sobre possíveis elementos de distração

Se você já discutiu dados, é provável que tenha encontrado correlações espúrias. Talvez você conheça o site que mostra correlações espúrias,¹ como a do exemplo a seguir:

Consumo de muçarela per capita e diplomas de Doutorado em Engenharia Civil concedidos

Correlação: 95,86% ($r=0,958648$)

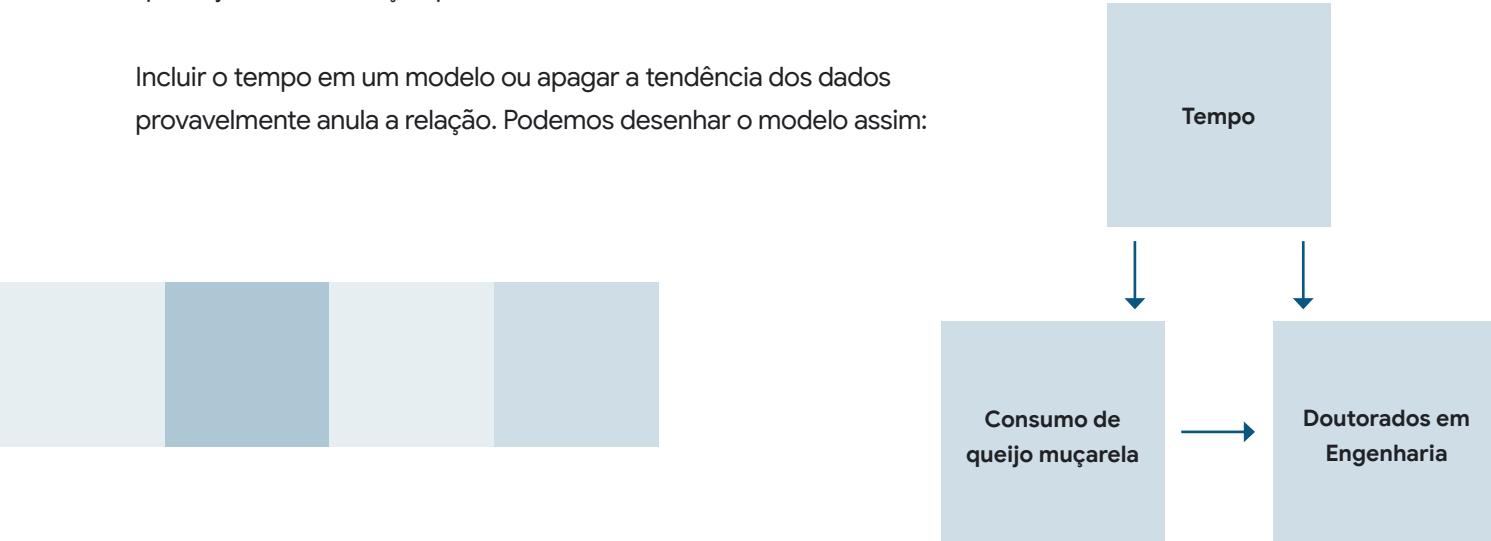


Fontes dos dados: Fundação Nacional da Ciência e Departamento da Agricultura dos EUA

—■— Consumo de queijo muçarela
—□— Doutorados em Engenharia

É pouco provável que exista uma relação causal entre doutorados em Engenharia e o consumo de queijo muçarela. Mesmo assim, há um elemento capcioso nessa relação: o tempo. Se o consumo de queijo muçarela e os doutorados em Engenharia estiverem com uma tendência positiva ao longo do mesmo período, é provável que haja uma correlação positiva.

Incluir o tempo em um modelo ou apagar a tendência dos dados provavelmente anula a relação. Podemos desenhar o modelo assim:



¹ <https://www.tylervigen.com/spurious-correlations>

Se não considerarmos o tempo (uma terceira variável), os dados podem apontar uma relação espúria entre o consumo de queijo muçarela e doutorados em engenharia.

Há ferramentas que ajudam pesquisadores a resolver isso, como o Dagitty (<https://dagitty.net/dags.html>). Essa ferramenta especifica nosso modelo causal. Para estimar adequadamente o efeito de X em Y, ele mostra as implicações do modelo e o que precisamos considerar ou *não*. Ferramentas como o Dagitty podem levar à conclusão de que a correlação não implica causalidade, mas quase sempre afeta o modo como a pessoa pensa sobre causalidade.

É impossível capturar todos os elementos que podem enviesar as estimativas de pesquisadores (imagine pesquisas de 50 horas e onisciência). Ainda assim, fazemos o possível para considerar os vieses, de modo que possamos chegar a estimativas precisas sobre os efeitos de diversas atividades, tecnologias e estruturas nos resultados que importam para você. Em última instância, muitos profissionais querem saber quais fatores afetam esses resultados principais. Modelos que não consideram os vieses não conseguem oferecer os direcionamentos necessários aos profissionais. Não queremos falar para alguém que queijo muçarela aumenta o desempenho de entrega de software, e esse erro é mais fácil de cometer do que você imagina.



Etapa 4. Desenvolver a pesquisa

O desenvolvimento da pesquisa envolve três aspectos: operacionalização, experiência e localização.

Operacionalização

Queremos medidas que capturem os conceitos em que estamos interessados de forma apropriada e confiável. Traduzir um conceito abstrato em algo mensurável é a arte da operacionalização. Essas medidas são os ingredientes que fundamentam toda a análise. Se nossas medidas não apontarem indícios claros, como vamos confiar no restante da análise? Como medimos um conceito tão vago quanto, por exemplo, produtividade? E esgotamento ou desempenho operacional?

Primeiro, buscamos na literatura se há medidas que já fazem isso com êxito. Se pudermos usar medidas que já foram validadas na nossa pesquisa, já temos uma ponte entre a pesquisa e toda a literatura relacionada a essa questão. O uso da *tipologia de culturas organizacionais* de Westrum é um exemplo de reutilização de medidas que já foram validadas antes.

No entanto, vários conceitos que nós pesquisamos ainda não foram validados. Nesse caso, fazemos uma pesquisa qualitativa para desvendar como as pessoas entendem o conceito e estudamos a literatura mais filosófica para entender os detalhes.

Experiência de pesquisa

Queremos que a pesquisa seja comprehensível, fácil, longa apenas o necessário e amplamente acessível. Essas metas são difíceis por causa do número de perguntas que precisamos fazer, do conhecimento técnico necessário para entender essas perguntas e das variações de nomenclatura para algumas práticas. Fazemos avaliações remotas e não moderadas para garantir que a pesquisa tenha um desempenho acima de alguns limites. Isso exige muitas iterações.

Localização

Pessoas do mundo todo participam da pesquisa anualmente. Este ano, trabalhamos para que a pesquisa fosse acessível para um público mais amplo. Por isso, ela foi localizada para inglês, espanhol, francês, português e japonês. Esse foi um esforço orgânico liderado por membros incríveis da comunidade da DORA. Além da contribuição de Googlers do mundo todo, também tivemos um parceiro dessa área: agradecemos a Zenika (<https://www.zenika.com>, em inglês) pela localização em francês. Esperamos que essas iniciativas sejam ampliadas para tornar a pesquisa verdadeiramente intercultural.

Etapa 5. Coletar respostas para a pesquisa

Usamos vários canais para recrutar participantes. Eles são divididos em duas abordagens: *orgânica* e *painel*.

A *abordagem orgânica* envolve usar todos os meios sociais disponíveis para comunicar que temos uma pesquisa para as pessoas responderem. Criamos postagens em blogs. Campanhas de e-mail. Postamos em redes sociais e pedimos para pessoas da comunidade fazerem o mesmo (ou seja, uma amostragem por bola de neve).

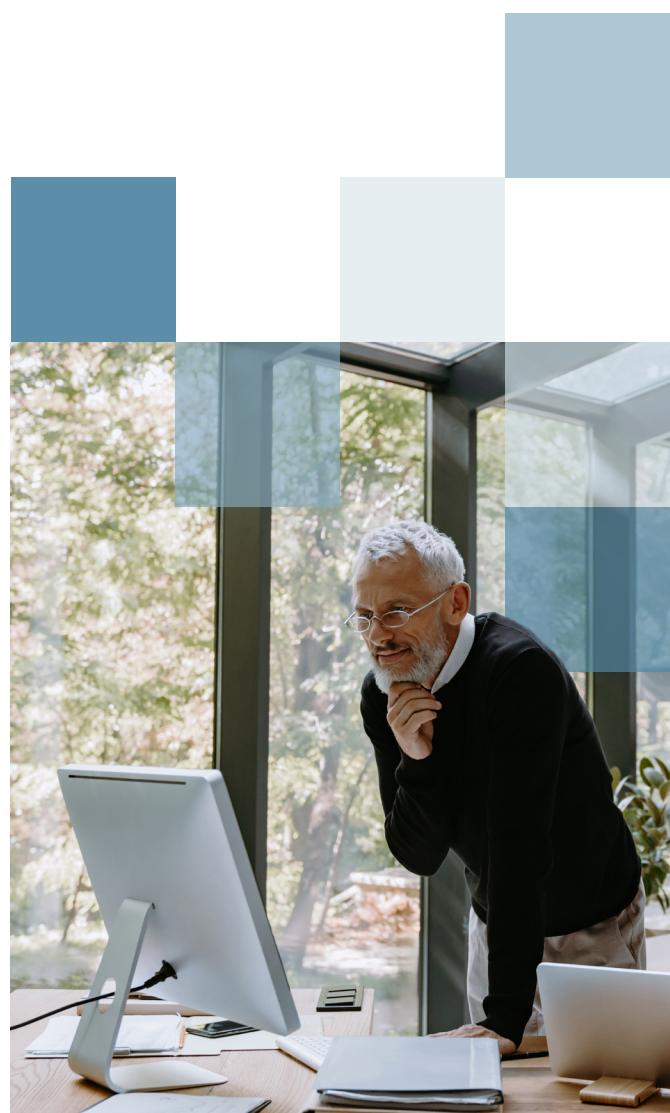
Usamos a *abordagem de painel* para complementar a orgânica. Aqui, tentamos recrutar pessoas tradicionalmente sub-representadas na comunidade técnica, além de buscar respostas de setores e tipos de organizações específicos. Em resumo, é aqui que temos algum controle sobre o recrutamento, o que não é possível na abordagem orgânica. A abordagem de painel também garante que vamos ter participantes suficientes, porque nunca sabemos se a abordagem orgânica vai gerar as respostas que precisamos para os tipos de análises que fazemos.

Etapa 6. Analisar os dados

A análise tem três etapas fundamentais: limpeza de dados, validação de medidas e avaliação de modelos.

Limpeza de dados

A limpeza de dados serve para aumentar a relação sinal-ruído. Por vários motivos, algumas respostas são ruídos. Alguns padrões de resposta podem indicar que a pessoa estava distraída, passando rápido pela pesquisa ou respondendo de má-fé. Isso é muito raro nos nossos dados, mas acontece. Respostas que não parecem representar a experiência real de uma pessoa são excluídas da análise porque são ruídos. Na hora de eliminar o ruído, o desafio é garantir que não estamos apagando o sinal, especialmente de forma tendenciosa ou que valide nossas hipóteses. Por exemplo, se concluirmos que ninguém poderia ter um valor alto em um quesito e baixo em outro, podemos acabar excluindo esse participante e alinhando os dados às nossas crenças, o que aumenta a probabilidade das nossas hipóteses serem confirmadas.



Validação de medidas

No início do relatório, mencionamos os conceitos que estamos tentando medir. Há várias terminologias que poderíamos adotar, mas uma delas é que a medida de um conceito é chamada de variável.² Essas variáveis são os ingredientes dos modelos, os elementos incluídos na pesquisa. Há duas maneiras amplas de analisar a validade dessas medidas: interna e externamente.

Para entender a validade interna da medida, analisamos o que nós achamos que indica a presença e um conceito. Por exemplo, um indicativo de uma documentação de qualidade é que ela é usada para resolver problemas.

A maioria das nossas variáveis é composta por vários indicadores, porque os construtos que nós queremos entender parecem ser multifacetados. Para entender essa natureza multifacetada de uma variável, testamos se os itens usados para representar esse construto combinam. Se sim, (ou seja, se compartilharem um alto nível de variância comunal), presumimos que há algo subjacente, como o conceito em que estamos interessados.

Pense na felicidade, por exemplo. A felicidade é multifacetada. Quando uma pessoa está feliz, esperamos que ela se sinta, aja e pense de uma determinada maneira. Presumimos que a felicidade está ligada a um padrão específico de sentimentos, ações e pensamentos. Portanto, esperamos que determinados sentimentos, ações e pensamentos surjam quando a felicidade está presente. Por isso, fariámos perguntas sobre esses sentimentos, ações e pensamentos. Nós usaríamos uma análise de fator confirmatório para testar se esses elementos surgem juntos.

Este ano, usamos o pacote lavaan do R para fazer essa análise.³ O lavaan retorna diversos dados estatísticos que ajudam a entender se um construto de fato representa como as pessoas respondem às perguntas. Se os indicadores de um conceito não combinarem, pode ser necessário revisar ou abandoná-lo, porque está claro que não encontramos uma forma confiável de medi-lo.

A validade externa de um construto envolve analisar como ele se encaixa no mundo. Podemos esperar que um construto tenha uma relação com outros construtos. Às vezes, espera-se que dois construtos tenham uma relação negativa, como felicidade e tristeza. Se nossa medida de felicidade tiver uma correlação positiva com a tristeza, podemos questionar nossa medida ou teoria. Da mesma forma, podemos esperar que dois construtos tenham relações positivas, mas não fortes. É provável que produtividade e satisfação no trabalho tenham uma correlação positiva, mas não idêntica. Se a correlação for alta demais, podemos dizer que é como se estivéssemos medindo a mesma coisa. Isso, por sua vez, significa que nossas medidas não estão calibradas o suficiente para captar as diferenças entre os dois conceitos ou que a diferença prevista na nossa hipótese não existe de fato.

² Moore, Will H. e David A. Siegel. A mathematics course for political and social research. Princeton University Press, 2013.

³ Rosseel, Y. "lavaan: An R Package for Structural Equation Modeling", Journal of Statistical Software, 48(2), 2012. 1–36. <https://doi.org/10.18637/jss.v048.i02>

Avaliação do modelo

Nas etapas 2 e 3, criamos modelos hipotéticos. Depois de limpar os dados na etapa 6, analisamos se esses modelos se encaixam neles. Este ano, adotamos uma abordagem bayesiana para entender a plausibilidade de várias hipóteses com base nos dados, em vez de tentar determinar a probabilidade dos dados com base em uma hipótese nula (ou seja, sem efeito presente). As principais ferramentas que usamos no R são blavaan⁴ e rstanarm.⁵ Assim, podemos testar se a probabilidade de um efeito é substancial ou dramática e não só pequena. Para avaliar um modelo, usamos a parcimônia. Isso envolve começar com um modelo simples e torná-lo mais complexo até que essa complexidade não seja mais justificável. Por exemplo, nós prevemos que a performance organizacional é o produto da interação entre o desempenho de entrega de software e o desempenho operacional.

Nosso modelo simples não inclui a interação:

Performance organizacional ~ Desempenho de entrega de software + Desempenho operacional

O segundo modelo incorpora a interação:

Performance organizacional ~ Desempenho de entrega de software + Desempenho operacional + Desempenho de entrega de software x Desempenho operacional

Seguindo as recomendações em “Regression and other stories”⁶ e “Statistical Rethinking”,⁷ usamos uma validação cruzada leave-one-out (LOOCV, na sigla em inglês) e o critério de informação amplamente aplicável de Watanabe–Akaike⁸ para determinar se a complexidade extra é necessária.

⁴ Merkle, Edgar C. e Yves Rosseel. “blavaan: Bayesian structural equation models via parameter expansion”, pré-impressão do arXiv, 2015. <https://doi.org/10.48550/arXiv.1511.05604>

⁵ Goodrich, Ben, Jonah Gabry, Imdad Ali e Sam Brilleman. “rstanarm: Bayesian applied regression modeling via Stan”. R package versão 2, nº 1 (2020).

⁶ Gelman, Andrew, Jennifer Hill e Aki Vehtari. Regression and Other Stories (Cambridge University Press, 2020).

⁷ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018.

⁸ Vehtari, Aki, Andrew Gelman e Jonah Gabry. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC”. Statistics and Computing, 27, 2017. 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>

Etapa 7. Relatar as descobertas

Depois, revisamos os resultados em equipe. Este ano, passamos alguns dias juntos em Boulder, Colorado, sintetizando os dados com a experiência de especialistas. Isso foi feito para cada capítulo do relatório, hipótese por hipótese. A interpretação de dados está sempre suscetível a riscos de enviesamento, especulação, anedotas e saltos lógicos. Esses riscos foram mitigados incluindo várias pessoas de origens diversas em uma sala que incentivou questionamento, divergência, perspectivas únicas e curiosidade.⁹

Depois de conseguir os resultados, os autores do relatório voltaram para seus respectivos cantos do mundo e começaram a escrever. Durante o processo de escrita, foram consultados editores e especialistas. Essas perspectivas foram fundamentais para comunicarmos nossas ideias. A pessoa responsável por analisar esses dados também garantiu que nenhuma afirmação contradisse o que os dados apontavam.

Esses capítulos foram condensados por nossos talentosos parceiros de design da BrightCarbon.¹⁰

Etapa 8. Sintetizar as descobertas com a comunidade

Contamos com o envolvimento da comunidade para encontrar maneiras de usar e interpretar as descobertas. Tentamos fazer recomendações específicas, mas no final há uma quantidade infinita de implementações que uma equipe pode testar com base nos resultados que encontramos. Por exemplo, a arquitetura com acoplamento flexível parece ser benéfica com base no que descobrimos. Mas certamente não há só uma forma de adotar uma arquitetura assim. Gerar e compartilhar abordagens com a comunidade é a única forma de aprimorar continuamente. Nossa mapa do mundo é uma interpretação e uma abstração do território e do contexto em que você, sua equipe e sua organização atuam.

Para participar da comunidade global de práticas da DORA, acesse o site <https://dora.community> (em inglês).

⁹ Stasser, G., & Titus, W. (1985). "Pooling of unshared information in group decision making: Biased information sampling during discussion". *Journal of Personality and Social Psychology*, 48(6), 1985. 1467–1478. <https://doi.org/10.1037/0022-3514.48.6.1467>

¹⁰ <https://www.brightcarbon.com/>

Informações demográficas e firmográficas

Quem respondeu à pesquisa

Há quase uma década, o programa DORA pesquisa capacidades, práticas e medidas de organizações de alto desempenho baseadas em tecnologia. Já ouvimos mais de 36.000 profissionais que trabalham em organizações de diversos tamanhos e setores.

Agradecemos por compartilhar seus insights! Este ano, quase 3.000 profissionais que trabalham em vários setores no mundo compartilharam experiências para ajudar a entender melhor os fatores que impulsionam organizações de alto desempenho baseadas em tecnologia.

As perguntas demográficas e firmográficas deste ano usaram pesquisas feitas pelo Stack Overflow. Mais de 70.000 pessoas participaram da pesquisa Stack Overflow Developer Survey de 2022.¹ Ela não atingiu todos os profissionais técnicos por uma série de motivos, mas é o mais próximo que temos de um censo do mundo de desenvolvedores. Tendo uma noção da população que participou dela, podemos encontrar vieses de resposta nos nossos dados e entender até que ponto nossas descobertas podem ser generalizadas. Esses dados e as perguntas demográficas e firmográficas feitas na Stack

Overflow Developer Survey foram bem elaboradas e valem a pena ser seguidas. Em relação à Stack Overflow Developer Survey, nosso conjunto de amostras tem uma proporção maior de mulheres, pessoas com deficiência e participantes que trabalham em organizações de grande porte. Ele é similar ao conjunto de amostras do Stack Overflow em questão de raça e etnia.

O número de participantes orgânicos deste ano foi 3,6 vezes maior do que o de 2022.

Na pesquisa deste ano, constatamos

3,6X

mais participantes orgânicos em comparação a 2022

¹ <https://survey.stackoverflow.co/2022#overview>

Informações demográficas

Gênero

Em relação a 2022, a amostra deste ano teve uma proporção menor de mulheres (12% em vez de 18%).

Gênero	% de participantes
Prefiro não dizer	3%
Ou, nas suas próprias palavras	2%
Mulher	12%
Homem	81%

Deficiência

Identificamos deficiências em seis dimensões que seguem as diretrizes do Washington Group Short Set.² Este é o quinto ano em que perguntamos sobre deficiências. A porcentagem de pessoas com deficiências diminuiu de 11% em 2022 para 6% em 2023.

Deficiência	% de participantes
Nenhuma das deficiências mencionadas	87%
Sim	6%
Prefiro não dizer/não respondeu	7%

Sub-representados

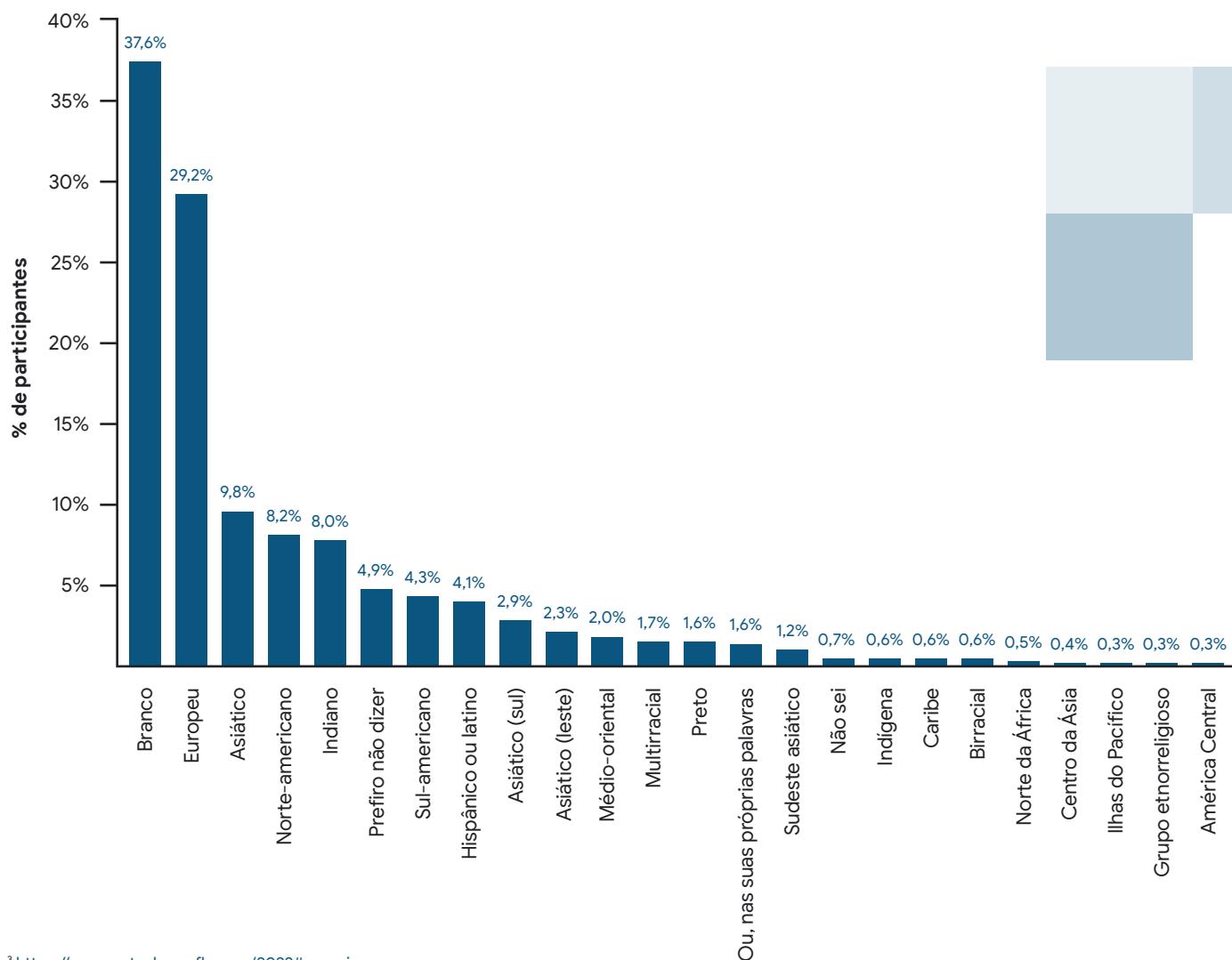
Identificar-se como membro de um grupo sub-representado pode se referir a raça, gênero ou outra característica. Este é o sexto ano em que perguntamos sobre a falta de representatividade. A porcentagem de pessoas que se identificam como sub-representadas diminuiu um pouco, de 19% em 2022 para 15% em 2023.

Sub-representados	% de participantes
Não	77%
Sim	15%
Prefiro não responder	7%

² <https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/>

Raça e etnia

Adotamos a pergunta da pesquisa com desenvolvedores do Stack Overflow de 2022.³ Como mencionado antes, nossa amostra é similar, mas com um desvio significativo: temos uma proporção menor de europeus.

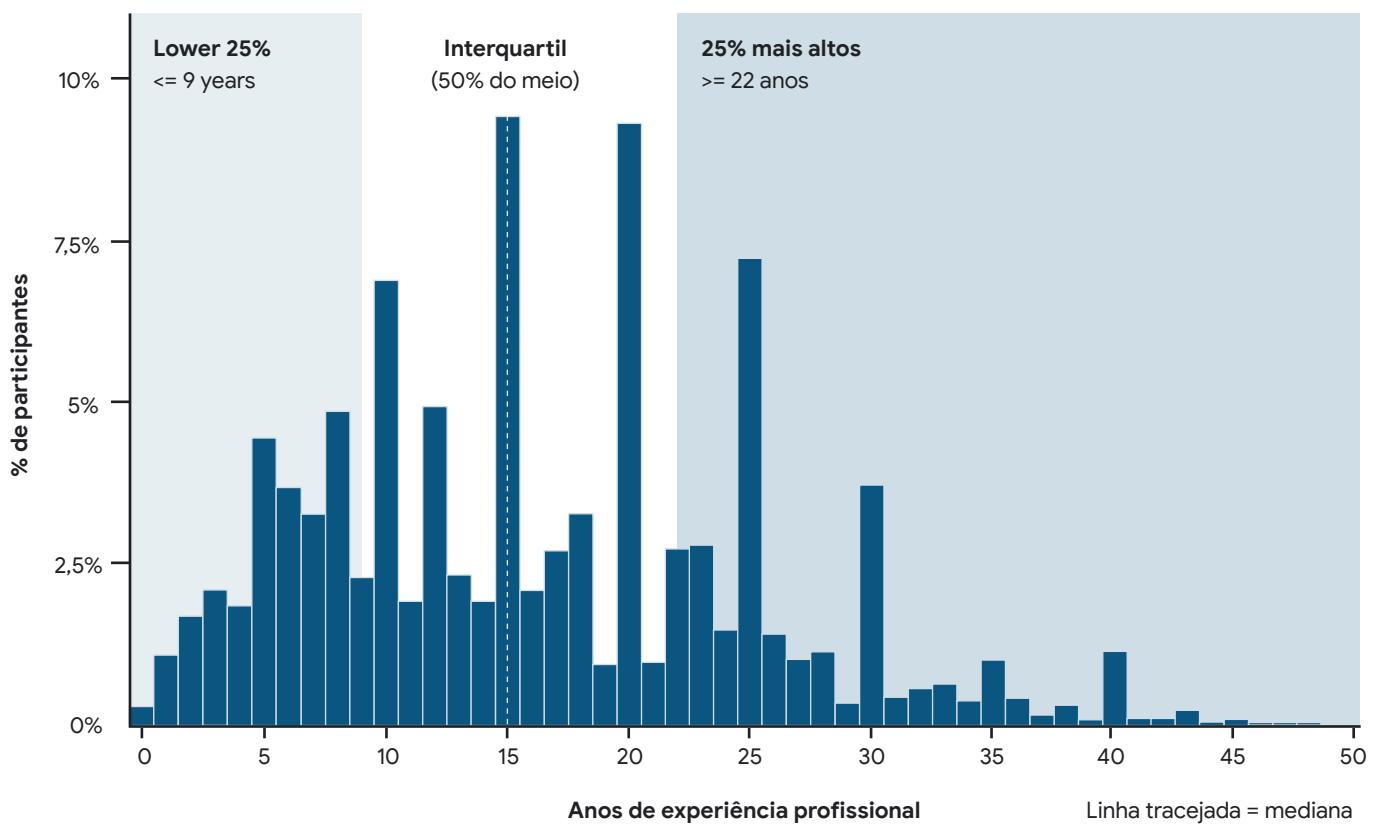


³ <https://survey.stackoverflow.co/2022#overview>

Anos de experiência

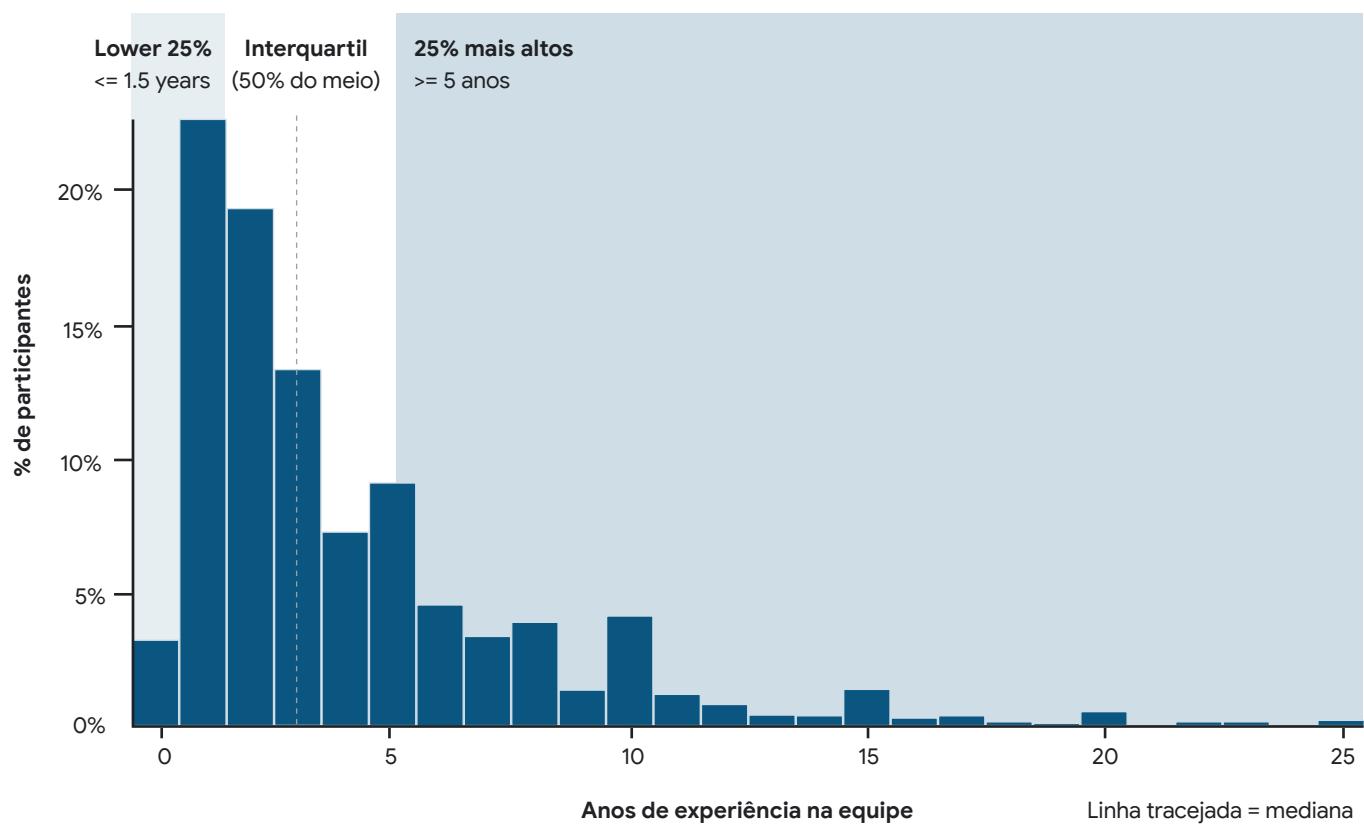
Experiência profissional

Queríamos entender há quanto tempo cada pessoa trabalha. Aqui, perguntamos “Quantos anos de experiência profissional você tem?”. No geral, coletamos dados de um grupo de profissionais um tanto experientes. Ou seja, 50% dos participantes tinham 15 anos de experiência ou mais. 25% dos participantes tinham mais de 22 anos de experiência. 25% dos participantes tinham menos de 9 anos de experiência. Em retrospectiva, não é muito óbvio o que cada pessoa entende como “trabalho” ou “profissional”.



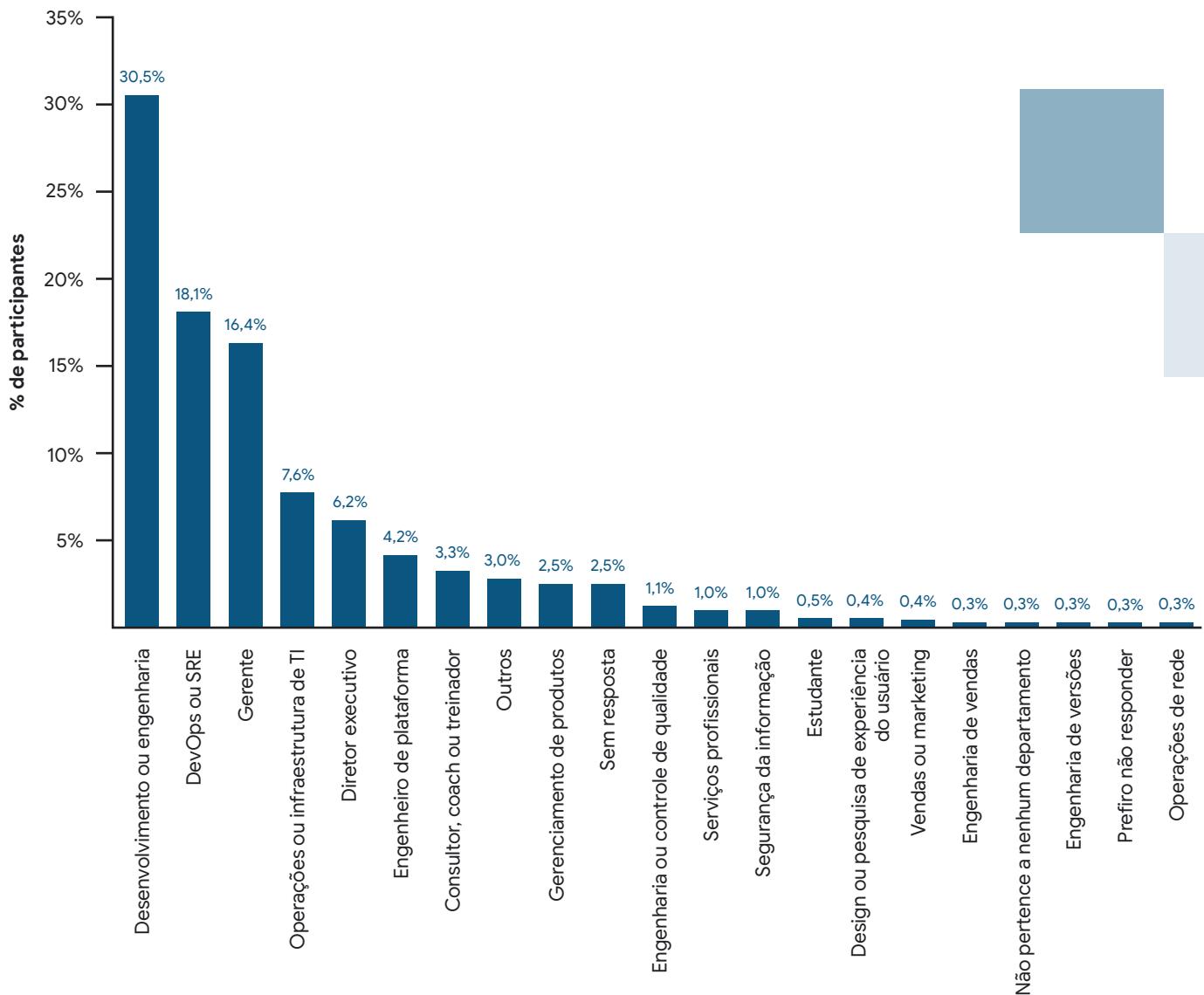
Experiência na equipe

Às vezes, trabalhar em uma nova equipe é como começar do zero. Nossa pergunta para analisar isso foi “Há quantos anos você trabalha na sua equipe atual?”. Apesar da alta experiência profissional, vários participantes são novos nas equipes. 50% dos participantes estão na nova equipe há menos de 3 anos. 25% estão na equipe há um ano e meio ou menos. Apenas 25% dos participantes estão na equipe há 5 anos ou mais. Será que isso reflete a mentalidade de aprimoramento contínuo dos participantes? Isso é resultado da instabilidade e do fluxo da economia?



Cargo

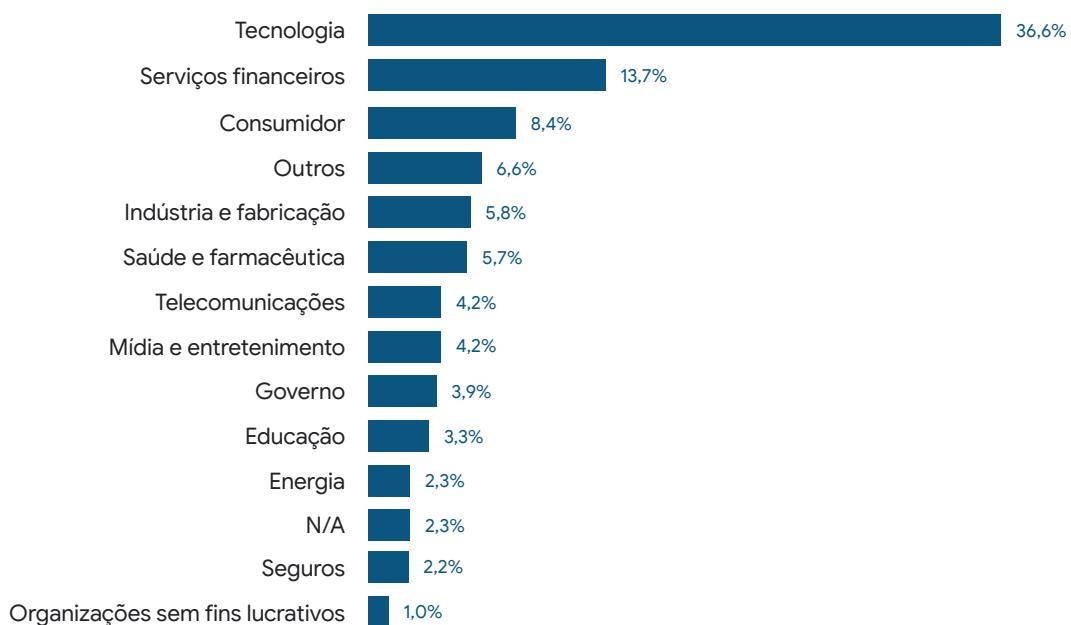
Dos participantes, 72% são pessoas que trabalham em equipes de desenvolvimento ou engenharia (30%), em equipes de DevOps ou de SRE (18%), em equipes de operações ou infraestrutura de TI (8%) ou são gerentes (16%). Em 2022, 85% dos participantes eram pessoas nesses cargos. A diminuição no número de participantes com um desses quatro cargos sugere que conseguimos atingir mais pessoas com outros cargos. A proporção de equipes de operações ou infraestrutura de TI (8%) voltou aos níveis de 2021 (9%) após o aumento súbito em 2022 (19%).



Firmografia

Setor

% de participantes



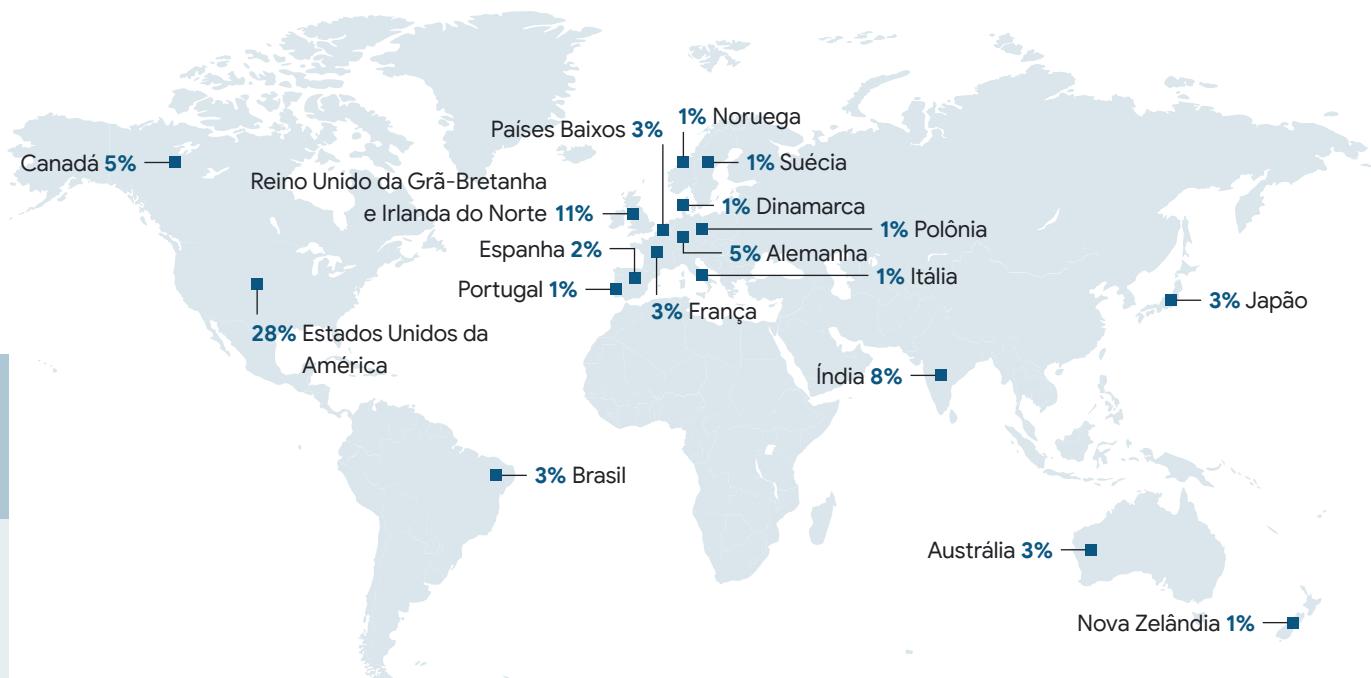
Número de funcionários

% de participantes



País

Sempre ficamos felizes em notar que pessoas do mundo inteiro participam da pesquisa. Agradecemos a todos!



País

EUA	Dinamarca	Lituânia	Tunisia	Bangladesh	Guatemala
Reino Unido	Suíça	Tailândia	Uruguai	República Dominicana	Honduras
Índia	Áustria	Hungria	Afeganistão	Gana	Letônia
Canadá	Quênia	Israel	Argélia	Hong Kong (RAE)	Libano
Alemanha	África do Sul	Vietnã	Egito	Cazaquistão	Luxemburgo
Austrália	Argentina	Emirados Árabes Unidos	Estônia	Mianmar	Maldivas
Brasil	República Tcheca	Bulgária	Islândia	Arábia Saudita	Malta
Não se aplica	Bélgica	Croácia	Irã	Somália	Maurício
Países Baixos	Colômbia	Equador	Nigéria	Sudão	Mongólia
Japão	Finlândia	Indonésia	Peru	Uganda	Marrocos
França	Irlanda	Filipinas	Eslováquia	Albânia	Nepal
Espanha	China	Armênia	Eslovênia	Bahamas	Catar
Suécia	Romênia	Georgia	Coreia do Sul	Bielorrússia	Antiga República Iugoslava da Macedônia
Itália	Singapura	Grécia	Sri Lanka	Bolívia	
Nova Zelândia	México	Malásia	Andorra	Camboja	Trinidad e Tobago
Polônia	Turquia	Paquistão	Angola	Costa Rica	República Unida da Tanzânia
Noruega	Ucrânia	Federação Russa	Antígua e Barbuda	Djibuti	
Portugal	Chile	Sérvia	Bahrein	El Salvador	Zimbábue

Formato de trabalho

Situação do emprego

88% dos participantes são funcionários em tempo integral. 10% dos participantes são prestadores de serviços. Alguns prestadores de serviços relatam ter experiências muito diferentes dos funcionários em tempo integral.

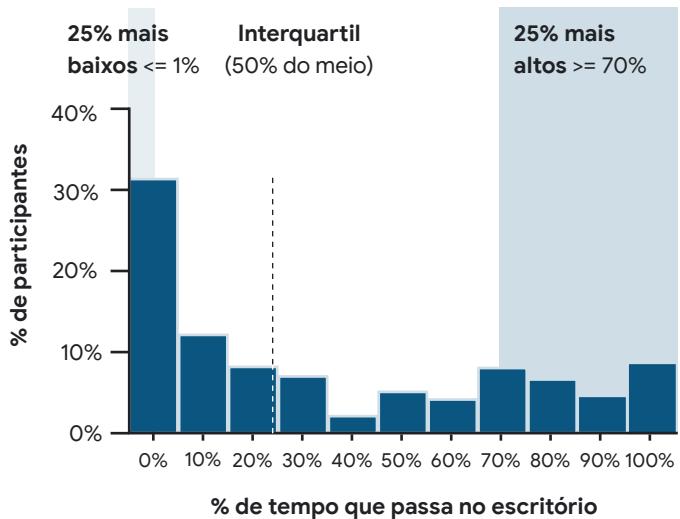
Contrato	% de participantes
Funcionário em tempo integral	88%
Prestador de serviço em tempo integral	8%
Funcionário de meio período	2%
Prestador de serviço de meio período	2%

*Pensando no aplicativo ou serviço principal em que você trabalha, qual é a melhor descrição do seu formato de trabalho com a organização que é proprietária desse aplicativo ou serviço?

Essa diferença de experiências pode ser resultado da posição que a pessoa ocupa na equipe. Alguns prestadores de serviço relatam que estão incorporados à equipe em que eles trabalham. Isso significa que a pessoa trabalha com os membros da equipe de perto todos os dias e considera a diferença entre ela e um funcionário em tempo integral irrelevante. 70% dos prestadores de serviço concordam total ou parcialmente com a afirmação de que estão incorporados à equipe.

Local

Este ano, o padrão de respostas indica que, apesar do esforço para voltar ao presencial, trabalhar de casa ainda é a realidade de muitos profissionais. Quase 33% dos participantes trabalham quase exclusivamente de casa (menos de 5% do tempo no escritório). 63% dos participantes trabalham mais de casa do que do escritório. Para os demais, o trabalho híbrido é o formato mais comum. Um indicativo disso é que 75% dos participantes passam menos de 70% do tempo no escritório. Não há muitas pessoas que tenham um vínculo forte com o escritório. Apenas 9% dos participantes passam mais de 95% do tempo no escritório.



Os modelos

Introdução

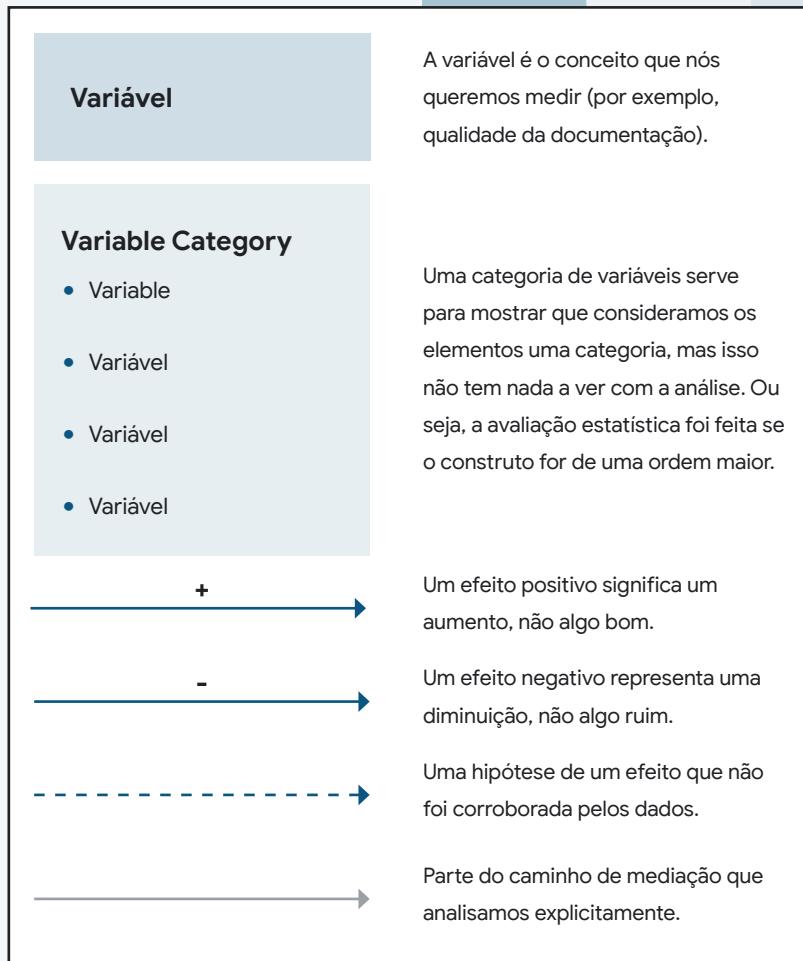
Nossa tradição era criar um modelo enorme. Este ano, decidimos separá-lo em vários modelos porque:

- Modelos enormes podem logo ficar complicados demais. Toda variável adicionada muda o funcionamento do modelo. Isso pode gerar estimativas imprecisas, além de ser mais difícil determinar por que uma mudança foi feita.
- Este ano, criamos hipóteses para cada uma das seções. Então, faz sentido criar um modelo para cada seção.
- Não existe um benefício claro de usar um modelo gigante para estimar o efeito de X em Y . Para entender o impacto de X em Y , usamos gráficos acíclicos dirigidos para definir quais covariáveis deveriam ser incluídas no modelo ou não.
- Devido ao número de hipóteses abordadas este ano, seria mais difícil entender um modelo gigante. Imagine como seria combinar todas as visualizações abaixo em uma só.



Como leio esses diagramas?

Depois que aprender a ler esses diagramas, você vai notar que são ferramentas eficientes para comunicar muitas informações.



Aviso: os modelos são resumos gerais!

Para facilitar a leitura, categorizamos algumas variáveis. Essa estratégia de categorização permite que o padrão geral dos resultados seja representado por uma seta que vai de ou para (ou tanto de quanto para) uma categoria de variáveis. No entanto, é possível que isso não seja verdadeiro para todas as variáveis da categoria. Por exemplo, o compartilhamento de conhecimentos tem um impacto positivo na maioria dos resultados principais. Portanto, seria traçada uma seta com o símbolo de positivo (+) do compartilhamento de conhecimentos para a categoria de variáveis que representa os resultados principais. No entanto, o compartilhamento de conhecimentos não tem um impacto positivo no desempenho de entrega de software. Para saber os detalhes, consulte os capítulos correspondentes.

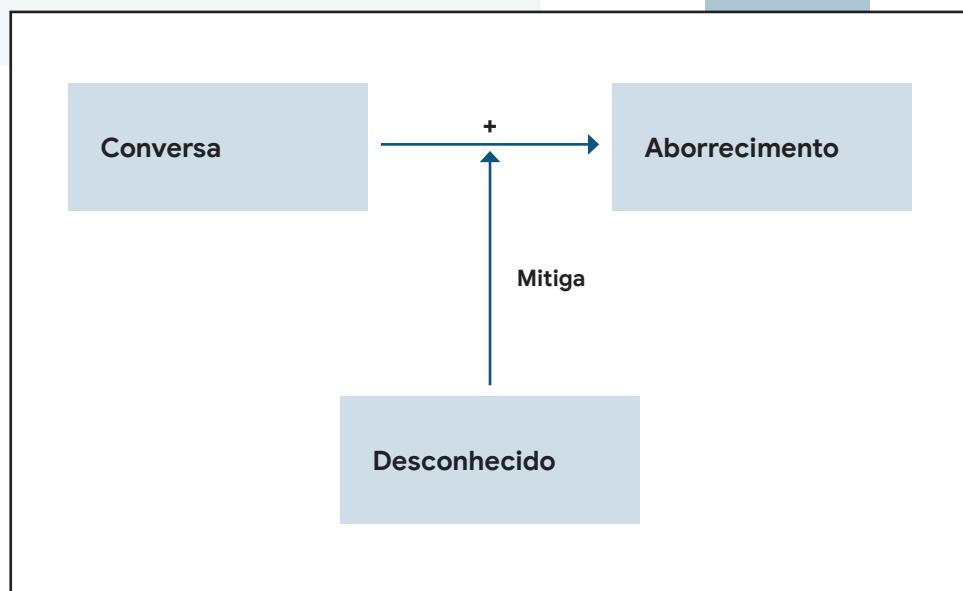
Exemplo de moderação

A moderação é um conceito complicado na estatística, mas no mundo real ela significa dizer “depende”. Vamos mostrar um exemplo para esclarecer o conceito de moderação no contexto deste relatório.

Na 3ª temporada de Curb Your Enthusiasm, Larry David fala: “Não gosto de falar com pessoas que eu conheço, mas não tenho dificuldade de falar com desconhecidos.” Com base nessa fala, podemos elaborar um diagrama para discutir:

“Não gosto de conversar com pessoas que eu conheço, mas não tenho dificuldade de falar com desconhecidos.”

Larry David



Este diagrama mostra que, para Larry, a conversa tem um impacto positivo no aborrecimento. Neste contexto, positivo significa aumento, não necessariamente algo bom. Isso é demonstrado pela linha preta entre conversa e aborrecimento, com a seta apontando para "aborrecimento". Essa seta sugere que há um fluxo causal da conversa para o aborrecimento. Pelas informações que temos, conversas costumam deixar Larry aborrecido.

A segunda constatação é que desconhecidos (que aqui são representados pelo booleano "desconhecido sim/não") não apontam para outra variável. Em vez disso, esse elemento aponta para um efeito, uma seta. Isso

significa que acreditamos que "desconhecido" não muda a variável, mas o efeito. Por isso a seta sai de "desconhecido" e aponta para outra seta, não outra variável. Estamos afirmando que, o efeito da conversa no aborrecimento é diferente se Larry estiver falando com um desconhecido ou não. Em outras palavras, estamos afirmando que o efeito da conversa no aborrecimento é diferente se a pessoa com quem Larry está conversando é um desconhecido ou não. Quando a pessoa é um desconhecido, o efeito da conversa é algo como Larry “não tem dificuldade com isso”. Podemos dizer que desconhecidos mitigam o aborrecimento que Larry sente quando conversa.

Há algumas formas de uma coisa moderar outra:

- **Intensificar**: tornar um efeito positivo mais positivo ou um efeito negativo mais negativo.
- **Atenuar**: enfraquecer o efeito.
- **Mitigar**: tornar efeitos positivos menos positivos e efeitos negativos menos negativos.
- **Reverter**: tornar efeitos positivos negativos e efeitos negativos positivos.
- **Modificar**: às vezes o efeito simplesmente muda, mas o padrão não pode ser explicado usando apenas uma palavra. Isso frequentemente acontece com variáveis categóricas como causas. Por exemplo, alguns setores têm comportamentos distintos dependendo das condições, como flutuações econômicas.

Exemplo de mediação

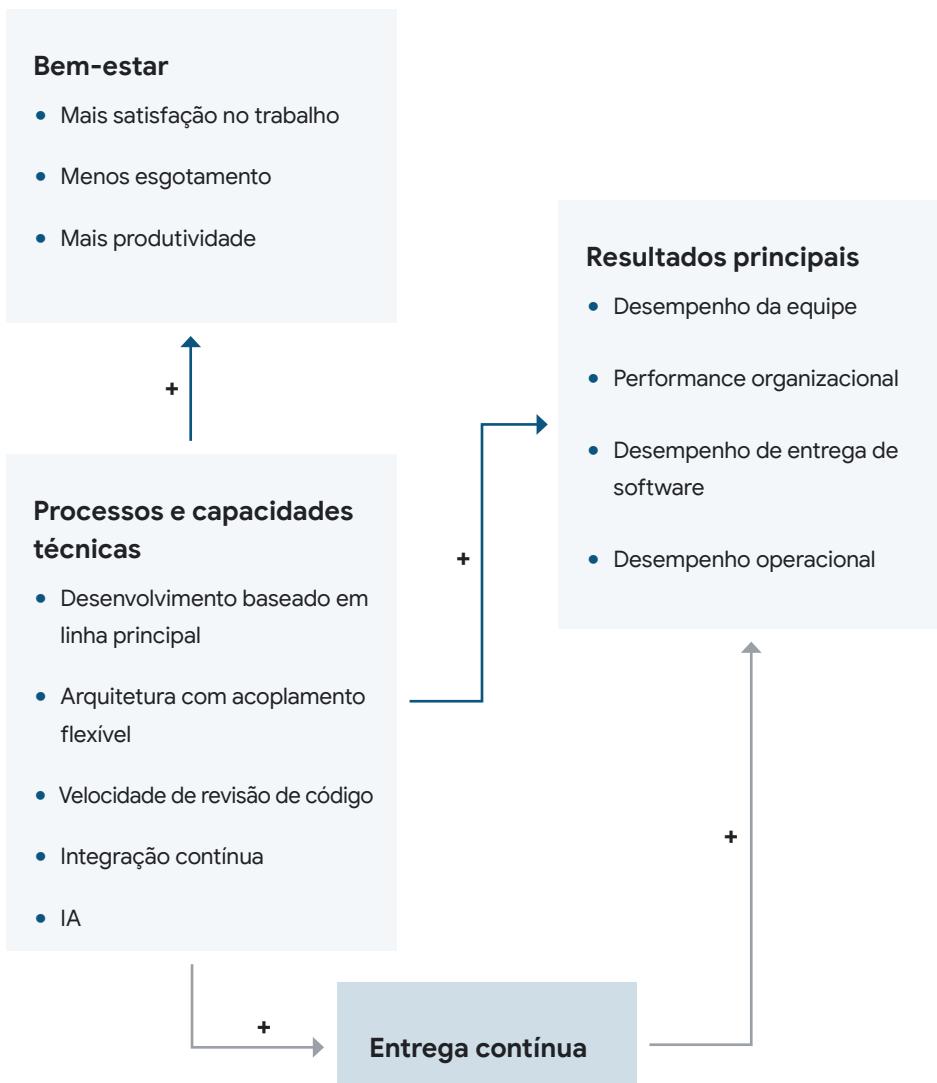
Além da moderação, também pensamos muito sobre a mediação. A raiz da mediação é o *porquê* ou o *como*. Por que o sol faz uma planta crescer mais? Por que o morango é doce? Como pisar no acelerador faz meu carro andar? Como analgésicos reduzem a dor? É possível testar a mediação usando dados estatísticos.¹ Isso significa que podem afirmar que “parece que o efeito de X em Y se deve total ou parcialmente a M”. Por exemplo, o efeito do sol no tamanho de uma planta se deve à fotossíntese.



¹<https://lavaan.ugent.be/tutorial/mediation.html>

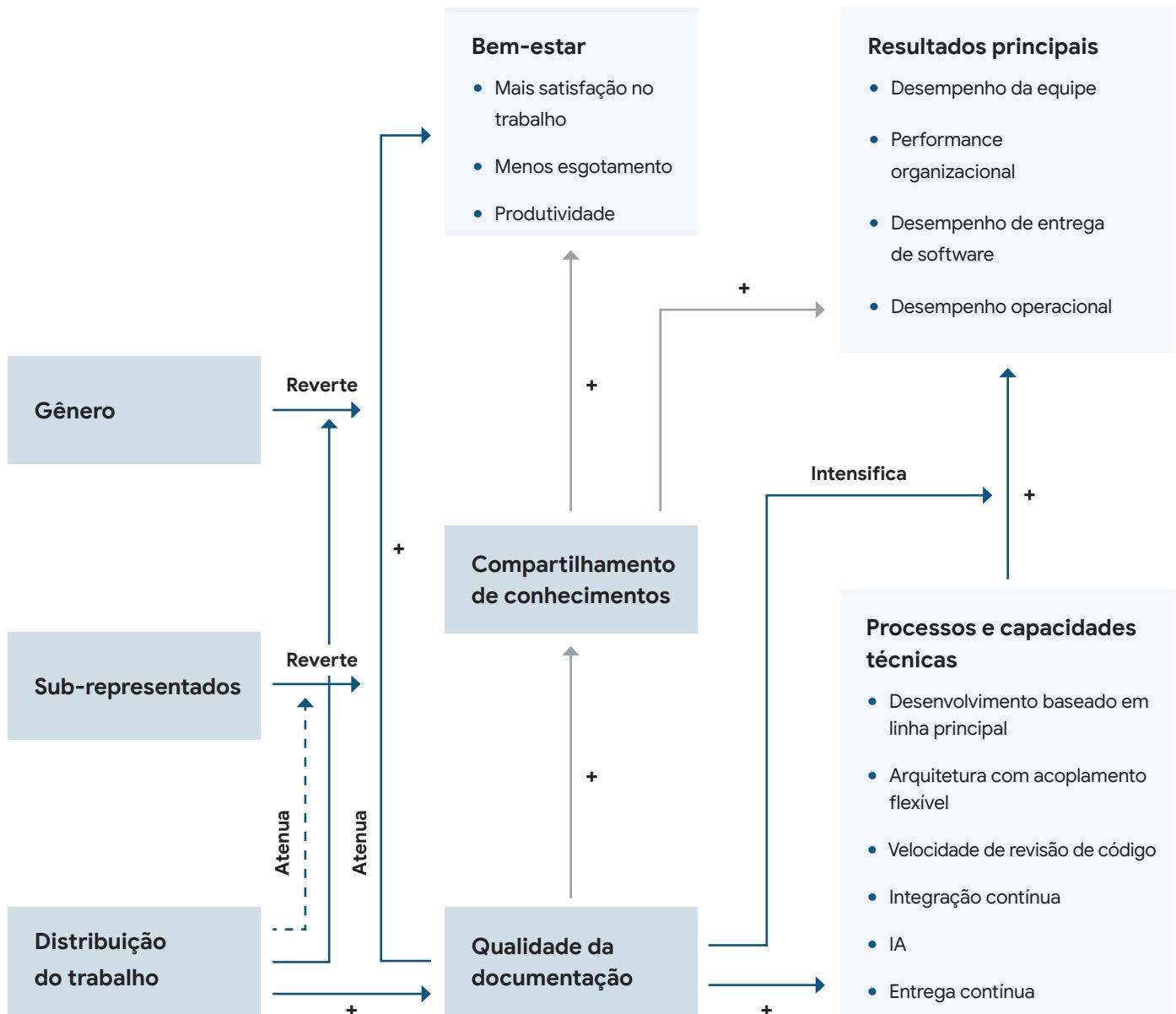
Modelo do Capítulo 3: Capacidades técnicas que predizem o desempenho

Processos e capacidades técnicas têm um impacto positivo no bem-estar e nos resultados principais. O efeito das capacidades técnicas nos resultados principais é parcialmente mediado pela entrega contínua, ao ponto de essas capacidades técnicas criarem um ambiente de entrega contínua que tem um efeito dominó nos resultados.



Modelo do Capítulo 4: O papel fundamental da documentação

Resumidamente, a documentação de alta qualidade tem efeitos positivos no bem-estar e nos resultados principais. Isso é parcialmente verdadeiro porque cria um ambiente que viabiliza o compartilhamento de conhecimento. A documentação de alta qualidade também ajuda as equipes a estabelecerem processos e capacidades técnicas. Além disso, ela ajuda processos e capacidades técnicas a terem um impacto ainda maior nos resultados principais. Por último, a qualidade da documentação não melhora o bem-estar de todos. Recomendamos ler a seção que explica essa descoberta complexa.



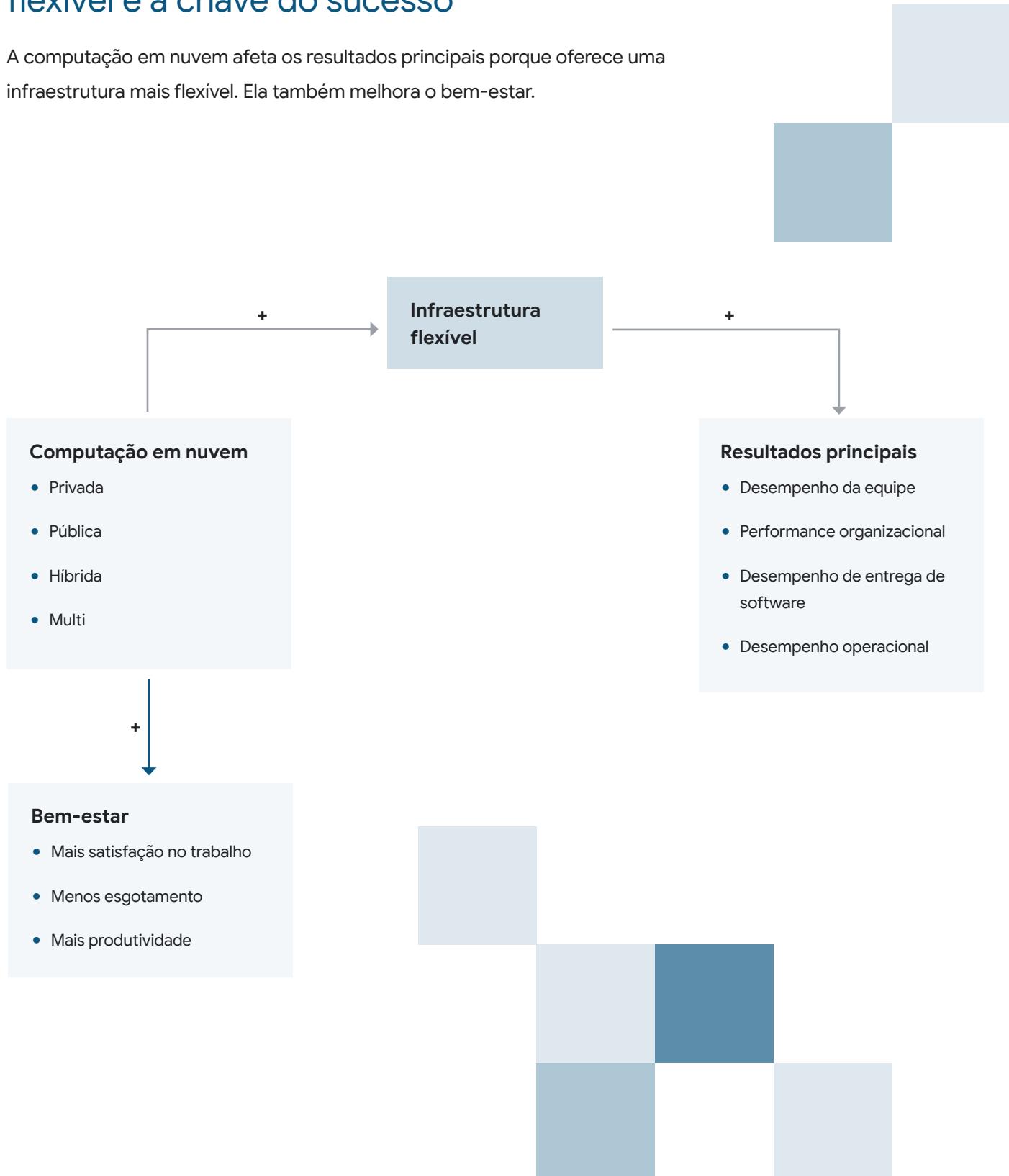
Modelo do Capítulo 5: A confiabilidade aumenta o desempenho

Aqui investigamos o papel central do desempenho operacional. Ele não apenas melhora o bem-estar e os resultados principais, mas também intensifica o efeito do desempenho de entrega de software. Práticas de confiabilidade têm uma relação não linear com o desempenho operacional. Recomendamos consultar esse capítulo para entender isso em mais detalhes.



Modelo do Capítulo 6: A infraestrutura flexível é a chave do sucesso

A computação em nuvem afeta os resultados principais porque oferece uma infraestrutura mais flexível. Ela também melhora o bem-estar.



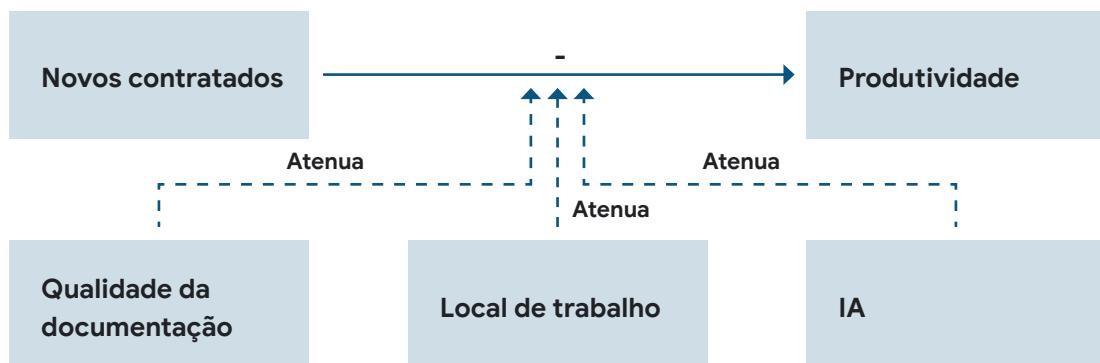
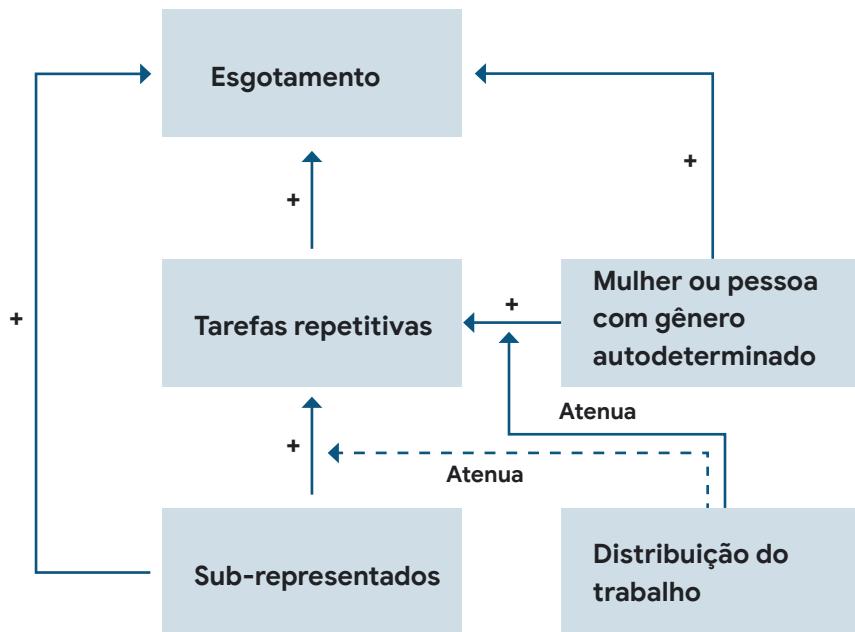
Modelo do Capítulo 7: Nada disso funciona sem investir na cultura

A cultura está no centro de boa parte deste diagrama. Descobrimos que a cultura tem uma relação positiva com as capacidades técnicas, os resultados principais e o bem-estar.



Modelos do Capítulo 8: A importância de "como", "quando", "por quê" e "quem"

Esta seção tem dois modelos. Um investiga por que e quando as pessoas que se identificam como sub-representadas e que não se identificam como homens sentem níveis mais altos de esgotamento. O outro investiga se a qualidade da documentação, o local de trabalho ou a IA pode ajudar novos contratados a serem mais produtivos.



Sugestões de leitura

Participe da Comunidade DORA para discutir, aprender e colaborar para melhorar o desempenho operacional e da entrega de software. [DORA.community](https://dora.community)

Faça a verificação rápida de DevOps da DORA.
<https://dora.dev/quickcheck>

Saiba as capacidades técnicas, processuais e culturais que impulsionam a entrega de software e a performance organizacional.
<https://dora.dev/devops-capabilities/>

Encontre recursos sobre SRE <https://sre.google> <https://goo.gle/enterprise-roadmap-sre>

Leia o livro: *Accelerate: The science behind devops: Building and scaling high performing technology organizations.* IT Revolution. <https://itrevolution.com/product/accelerate/>

Descubra um conjunto adequado e métricas para sua equipe usando o framework SPACE. “The SPACE of Developer Productivity: There’s more to it than you think”. <https://queue.acm.org/detail.cfm?id=3454124>

Há várias pesquisas sobre revisões de código modernas. Aqui estão alguns relatórios para você conhecer:

- “Expectations, Outcomes, and Challenges of Modern Code Review” <https://dl.acm.org/doi/10.5555/2486788.2486882>

- “Code Reviews - From bottlenecks to Superpowers” <https://learning.acm.org/techtalks/codereviews>
- “Modern Code Review- A Case Study at Google” <https://dl.acm.org/doi/10.1145/3183519.3183525>
- “The Choice of Code Review Process: A Survey on the State of the Practice” https://link.springer.com/chapter/10.1007/978-3-319-69926-4_9
- “Investigating the effectiveness of peer code review in distributed software development based on objective and subjective data” <https://jserd.springeropen.com/articles/10.1186/s40411-018-0058-0>

Leia o livro: *The No Club: Putting a Stop to Women’s Dead-End Work.* Simon & Schuster. <https://www.simonandschuster.com/books/The-No-Club/Linda-Babcock/9781982152338>

Publicações do programa de pesquisa da DORA, incluindo relatórios anteriores do Accelerate State of DevOps. <https://dora.dev/publications/>

Perguntas frequentes sobre a pesquisa e os relatórios. <http://dora.dev/faq>

Erratas, mudanças, correções e esclarecimentos sobre este relatório podem ser enviados para <https://dora.dev/publications/errata>

Apêndice

Refinamento das medidas de desempenho de entrega de software

Este ano, decidimos mudar a forma de avaliar falhas nas mudanças e como as pessoas se recuperam delas.

Mudamos a forma dos participantes relatarem a **taxa de falha nas mudanças**. Nos anos anteriores, os participantes recebiam seis opções (0-15%, 16-30% etc.). Este ano, usamos uma barra deslizante para que pudessem escolher qualquer valor entre 0% e 100%.

Isso foi feito por dois motivos:

- A taxa de falha nas mudanças sempre se comportou de um jeito diferente em comparação às outras três medidas de desempenho de entrega de software.¹ Nossa teoria é que isso se deve em parte ao tamanho dos conjuntos. Se as respostas forem mais precisas, talvez a resposta seja melhor para fins estatísticos. Essa teoria foi validada.
- Temos a hipótese de que as equipes entendem melhor a própria taxa de falha nas mudanças hoje do que quando a pesquisa começou, há quase uma década. Não temos uma forma de validar essa hipótese.

Este ano, usamos duas perguntas sobre recuperação de falhas:

- Antes, a pergunta usada era esta: “Pensando no aplicativo ou serviço principal em que você trabalha, quanto tempo geralmente leva para restaurar o serviço depois que ocorre um incidente ou um defeito que afeta os usuários (por exemplo, interrupção não planejada ou obstáculo no serviço)?”
- Este ano, adicionamos os seguintes qualificadores (as diferenças estão em negrito aqui, mas não estavam assim na pesquisa): “Pensando no aplicativo ou serviço principal em que você trabalha, quanto tempo geralmente leva para restaurar o serviço **depois que uma alteração na produção ou a liberação para os usuários resulta em queda no serviço (por exemplo, leva à deficiência ou interrupção do serviço)** e, **posteriormente, exige remediação (como um hotfix, reversão, correção de encaminhamento, patch)**? ”

¹ Forsgren, N., Humble, J., and Kim, G. *Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations* (IT Revolution Press, 2018), 37–38.

A forma anterior de perguntar sobre tempo de recuperação não permitiu diferenciar uma falha gerada por uma mudança no software de uma causada por um terremoto que interrompeu o serviço de um data center. Tínhamos a hipótese de que usar uma pergunta mais precisa nos permitiria comparar falhas semelhantes e mais estatisticamente alinhadas às outras três medidas de desempenho de entrega de software.

Agora estamos usando o termo “tempo de recuperação após falha ao implantar” para diferenciar nossa medida do termo mais genérico “tempo de restauração” que usávamos antes e às vezes é abreviado como “MTTR” (sigla em inglês para “tempo médio de reparo”). O MTTR gerou mais confusão na comunidade: o “M” é de médio ou mediano? Além disso, os profissionais que querem saber mais sobre falhas, como as do espaço de engenharia de resiliência, estão deixando de usar o MTTR como medida confiável para guiar o aprendizado e o aprimoramento.²

A pergunta recém-adicionada e a nova métrica, o tempo de recuperação após falha ao implantar, estão mais alinhadas ao objetivo de medir o desempenho de entrega de software.

² “Moving Past Simple Incident Metrics: Courtney Nash on the VOID”
<https://www.infoq.com/articles/incident-metrics-void/>



A matemática por trás das comparações

Ao longo deste relatório, há algumas estatísticas que indicam que um valor mais alto para uma determinada variável faz uma pontuação específica ficar N vezes maior. O que é isso? Ao que isso está relacionado? Aqui está nossa receita para criar essas fórmulas:

1. Para criar uma fórmula baseada no modelo avaliado, use técnicas de regressão que consideram possíveis vieses:

$$\text{Felicidade} \sim 5,64 + 0,19 * \text{raio de sol} + 0,14 * \text{temperatura}$$

2. Para a variável de interesse, como o raio de sol, encontre valores altos e baixos para comparar:

$$\text{média} = 6,3; \text{dp} = 1,4$$

$$\text{Nível alto de raio de sol} = 1 \text{ dp acima da média} = \text{média} + \text{dp} = 6,3 + 1,4 = 7,7$$

$$\text{Nível baixo de raio de sol} = 1 \text{ dp abaixo da média} = \text{média} - \text{dp} = 6,3 - 1,4 = 4,9$$

3. Calcule a média das covariáveis. Neste exemplo, isso seria a temperatura, que é igual a 0,24 (temperatura padronizada).
4. Preencha a fórmula da primeira etapa com valores altos e baixos para raio de sol. Você vai notar que apenas um número da fórmula é diferente. É assim que mantemos todo o restante igual e isolamos apenas aquilo em que estamos interessados.

Felicidade prevista com alto nível de raio de sol = $5,64 + 0,19 * 7,7 + 0,14 * 0,24 = 7,1$

Felicidade prevista com baixo nível de raio de sol = $5,64 + 0,19 * 4,9 + 0,14 * 0,24 = 6,6$

5. Calcule a proporção:

$$\frac{\text{felicidade prevista com alto nível de raio de sol}}{\text{felicidade prevista com baixo nível de raio de sol}} = \frac{7,1}{6,6} = 1,1x$$

6. Essa proporção sugere que altos níveis de raio de sol geram níveis de felicidade 10% mais altos em comparação a baixos níveis de raio de sol.

O que é uma “simulação”?

Isso não significa que os dados são fictícios. Usamos estatísticas bayesianas para calcular uma **posterior**, que tenta capturar a “frequência esperada em que diferentes valores de parâmetros vão aparecer.”³ A “simulação” significa calcular uma posterior mais de 1.000 vezes para investigar quais valores são mais confiáveis para uma parâmetro (média, peso beta, sigma, interceptação etc.) para nossos dados. “Imagine que a posterior é um balde cheio de valores de parâmetros, números como 0,1, 0,7, 0,5, 1 etc. Dentro do balde, cada valor existe em proporção a sua própria probabilidade posterior, de modo que valores próximos do topo são muito mais comuns que os mais baixos.”⁴

Isso tudo para dizer que usamos simulações para investigar possíveis interpretações dos dados e ter uma noção de quanta incerteza existe nelas. Pense em cada

simulação como uma pequena IA que só conhece nossos dados e tem algumas regras para preencher uma lacuna (o parâmetro) com um palpite baseado nessas informações. Isso é feito 4.000 e você recebe palpites de 4.000 pequenas IAs para um determinado parâmetro. Esses palpites podem ensinar muitas coisas. Por exemplo, qual é a média dos palpites, entre quais valores 89% dos palpites ficam, quantos palpites estão acima de um determinado nível, quanto esses palpites são variados etc. É possível até combinar palpites (simulações) de vários modelos.

Quando mostramos um gráfico com várias linhas ou uma distribuição de possíveis valores, estamos tentando mostrar o que é mais plausível com base nos nossos dados e quanta incerteza existe ali.

³ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018, pg. 50

⁴ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018, pg. 52

