

Все что мы узнали о Data Science в России

В декабре 2017 года мы провели опрос среди различных групп пользователей, так или иначе имеющих отношение к анализу данных. Мы распространяли ссылку на опрос в сообществах Slack OpenData Science, рассылали напрямую в компании, в которых есть подразделения data analysis, а также участникам конференции SmartData. Нам очень хотелось узнать, какими языками программирования, технологиями и инструментами пользуются специалисты в этой области.

Здесь мы делимся результатами этого исследования. При всем желании мы бы не смогли отразить все интересные наблюдения, которые можно сделать из полученных данных. Поэтому если вы хотите поработать с исходными данными, можете скачать их [здесь](#). В целях конфиденциальности все ответы на открытые вопросы удалены.

Оглавление

1. Профиль специалиста в области Data Science
2. Языки программирования
3. Технологии и инструменты
4. Деньги

1. Профиль специалиста в области Data Science

57% респондентов относятся к возрастной категории от 21 до 29 лет, возраст 35% опрошенных — от 30 до 39. Это вполне соответствует ситуации в мире, где такое соотношение составляет 48% и 29%, согласно нашему исследованию Data Science Survey Worldwide*. Что интересно, доля тех, кому за 40, превосходит долю тех, кому меньше 20 лет.

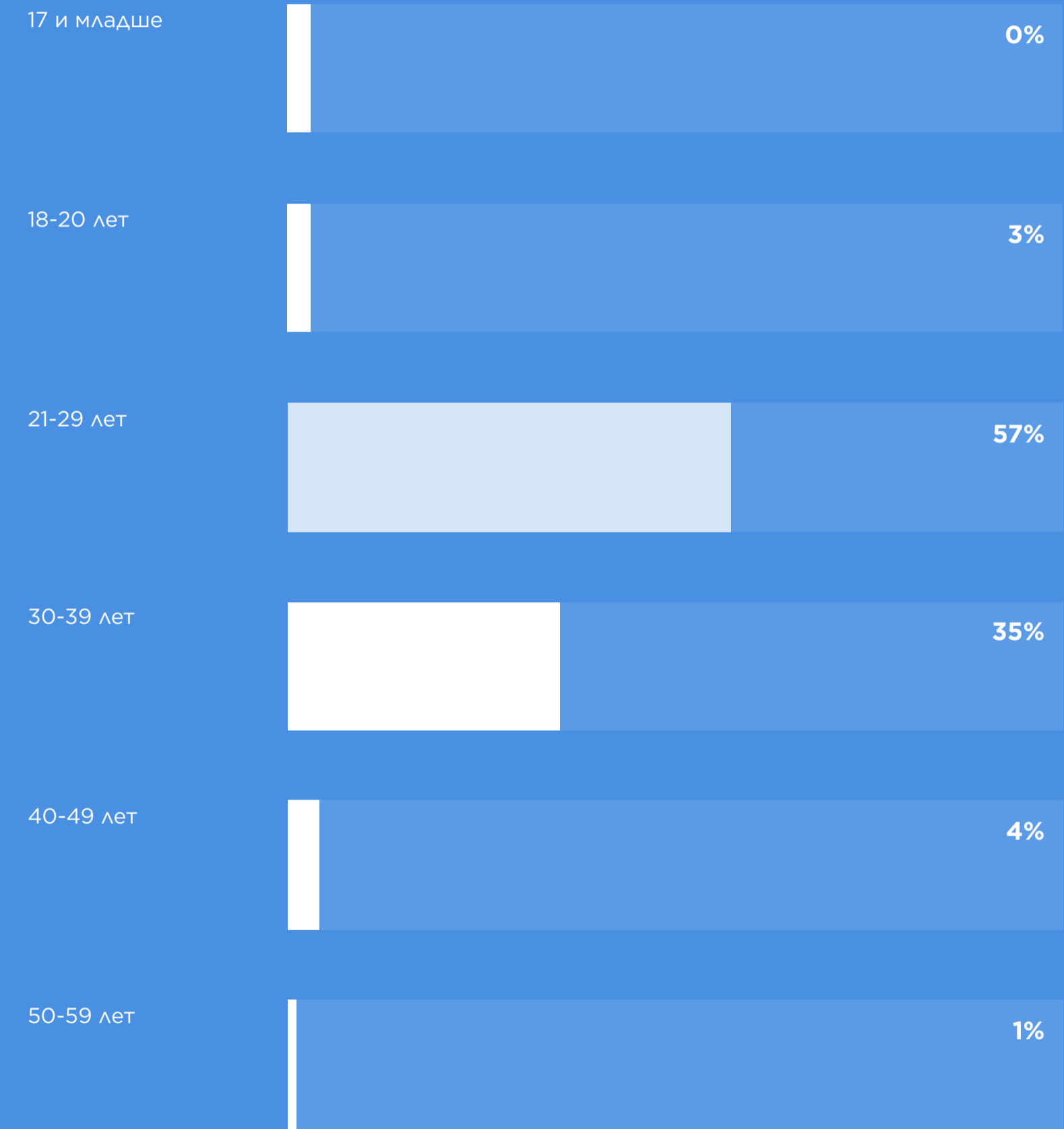


Юрий Харчев
заместитель руководителя Capital Markets,
Райффайзен Банк, Москва

«Исследование JetBrains полезно тем, что просто и понятно описывает существующий ландшафт индустрии анализа данных в России. Выводы могут оказаться полезны как для процесса формирования data science команд, так и при поиске отдельных экспертов».

*Data Science Worldwide — опрос JetBrains, который проводился весной 2018 и по списку вопросов почти полностью соответствует опросу «Все, что мы хотели узнать о Data Science в России, но боялись спросить». Результаты DS Worldwide будут опубликованы отдельно осенью 2018 года.

Сколько Вам лет?

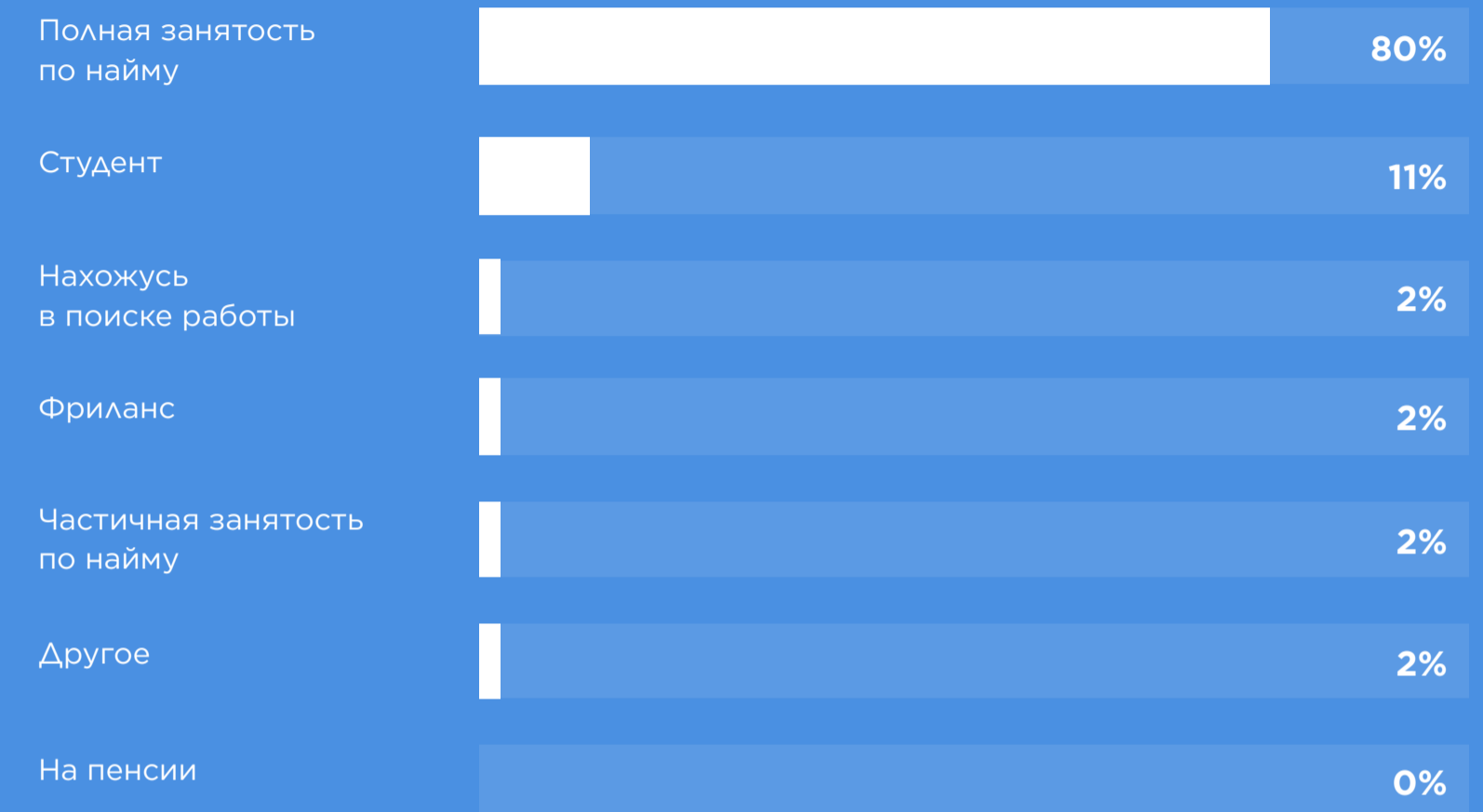


1. Профиль специалиста в области Data Science



специалистов трудоустроены и имеют полную занятость

Каким образом вы в основном трудоустроены?



Укажите уровень вашей квалификации



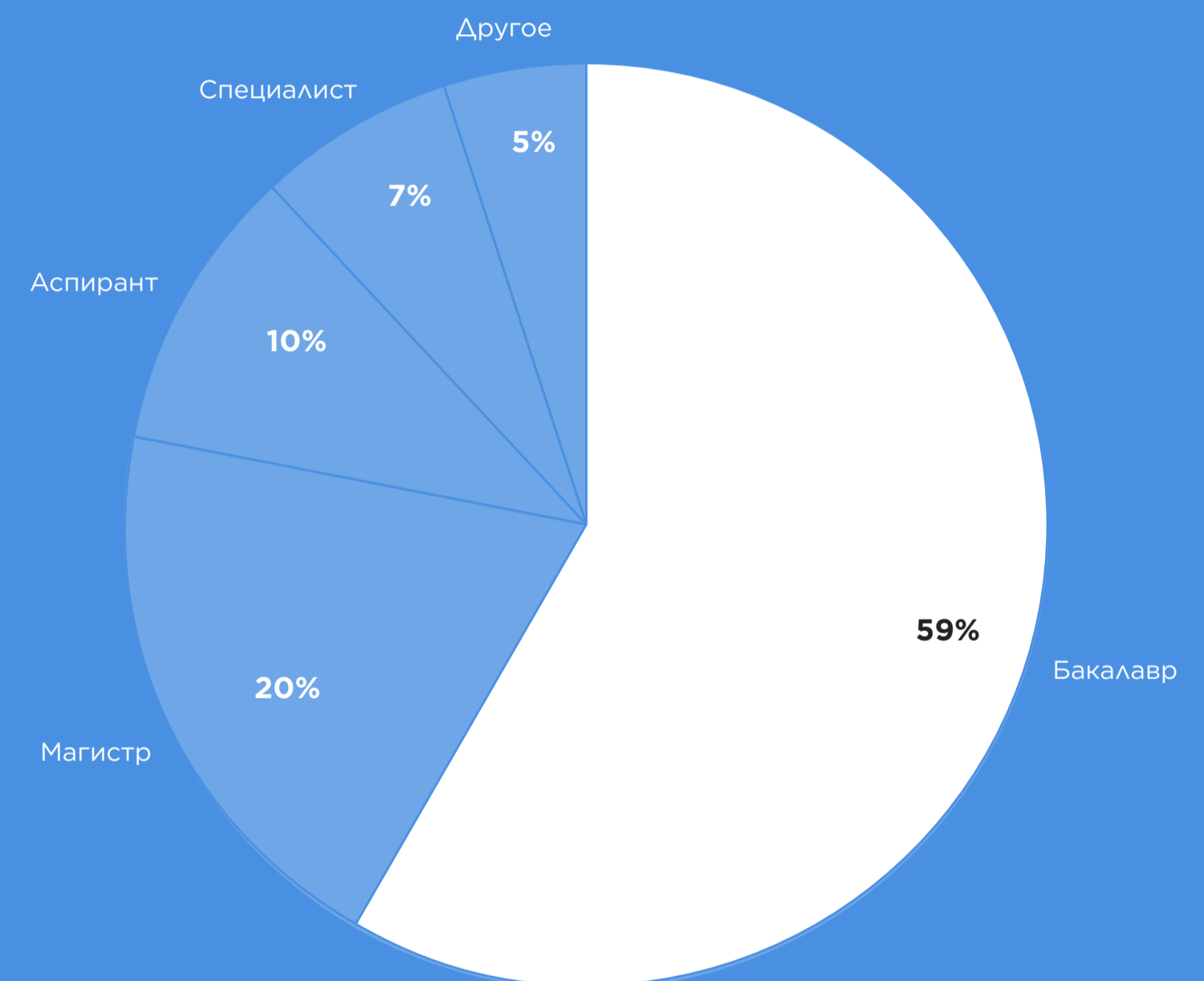
1. Профиль специалиста в области Data Science

Однако распределение респондентов по уровню образования сильно отличается от данных по миру.

Среди опрошенных 59% имеют степень бакалавра, и только 20% степень магистра, в то время как в мире согласно Kaggle, степень бакалавра имеют 32% опрошенных, а степень магистра — 41%.

Таким образом среди респондентов уровень образования ниже, чем в мире, при том что возрастная группа в целом совпадает.

Укажите полученное вами образование



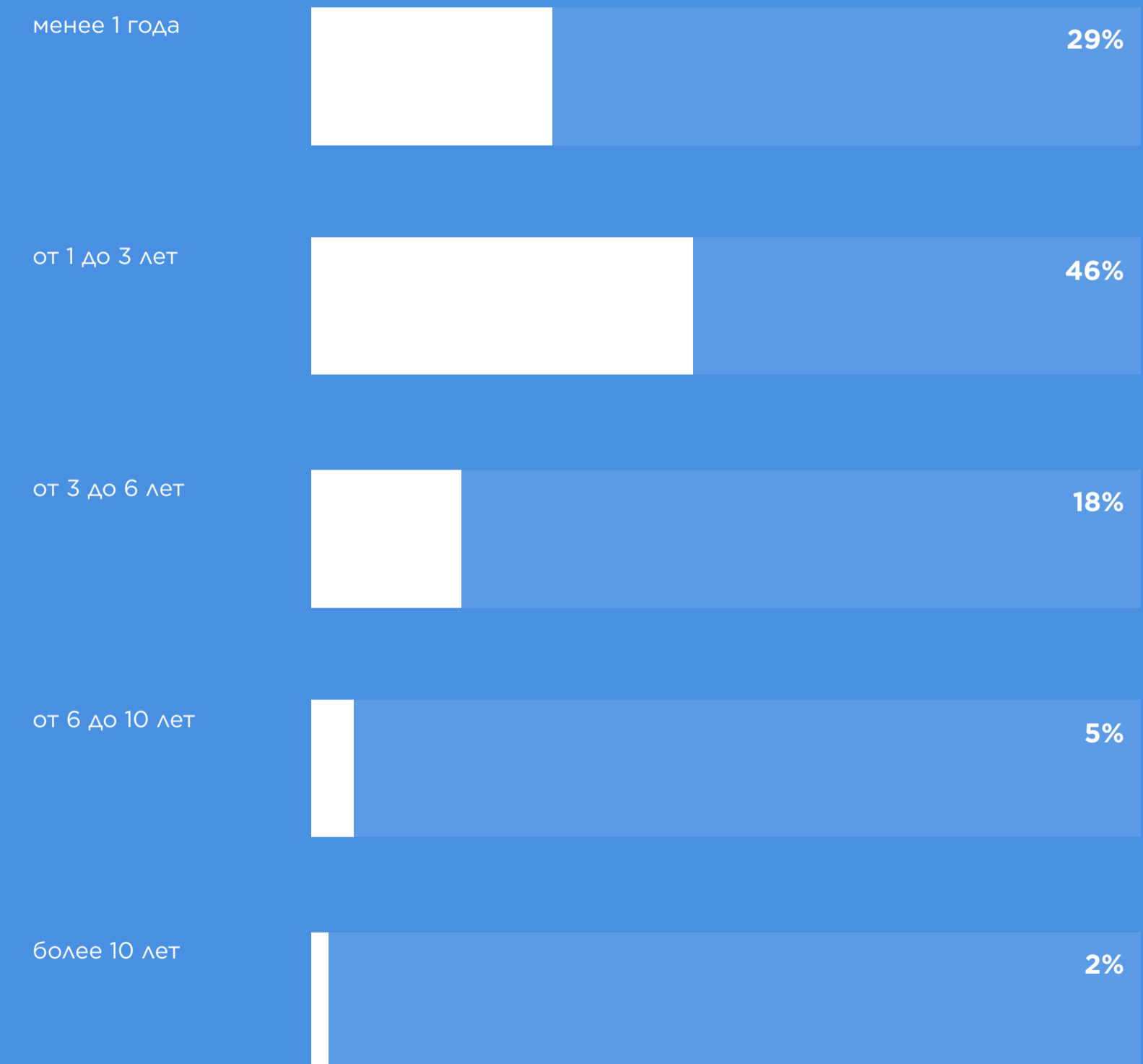
1. Профиль специалиста в области Data Science

Как можно увидеть, около половины респондентов (46%) работают в области анализа данных от 1 до 3 лет. И лишь 18% имеют опыт 3-6 лет.

Необычно, что для абсолютного большинства респондентов (тех, у кого опыт от 0 до 6 лет, а таких >90%) средний возраст никак не связан с опытом.

То же самое относится и к занимаемой позиции: только возраст старших специалистов (seniors) несколько превышает возраст всех остальных.

Как долго Вы работаете в сфере анализа данных?



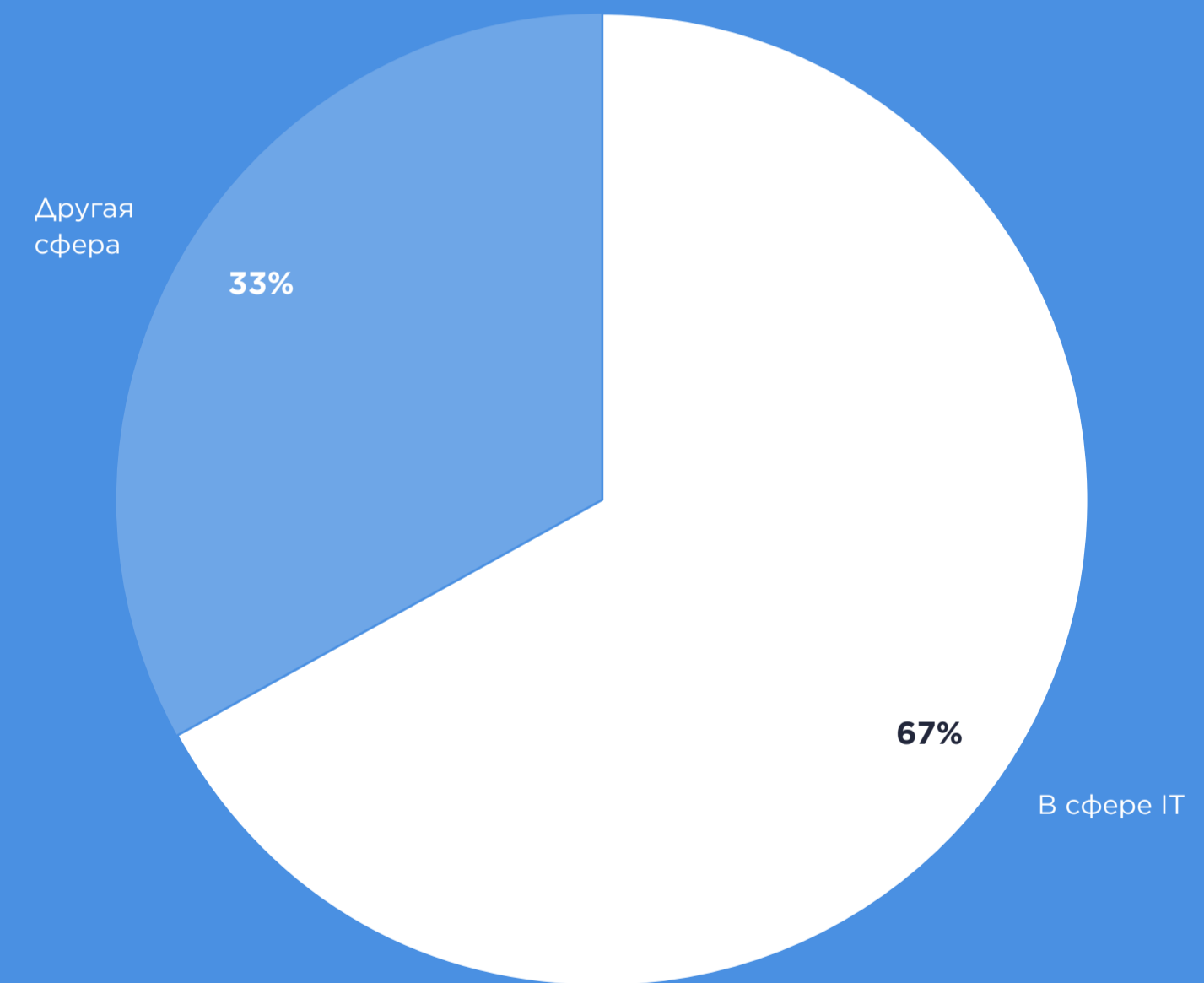
1. Профиль специалиста в области Data Science

Сфера информационных технологий является основной областью применения анализа данных.



опрошенных работают в научной сфере

В какой области вы в основном применяете анализ данных?



1. Профиль специалиста в области Data Science

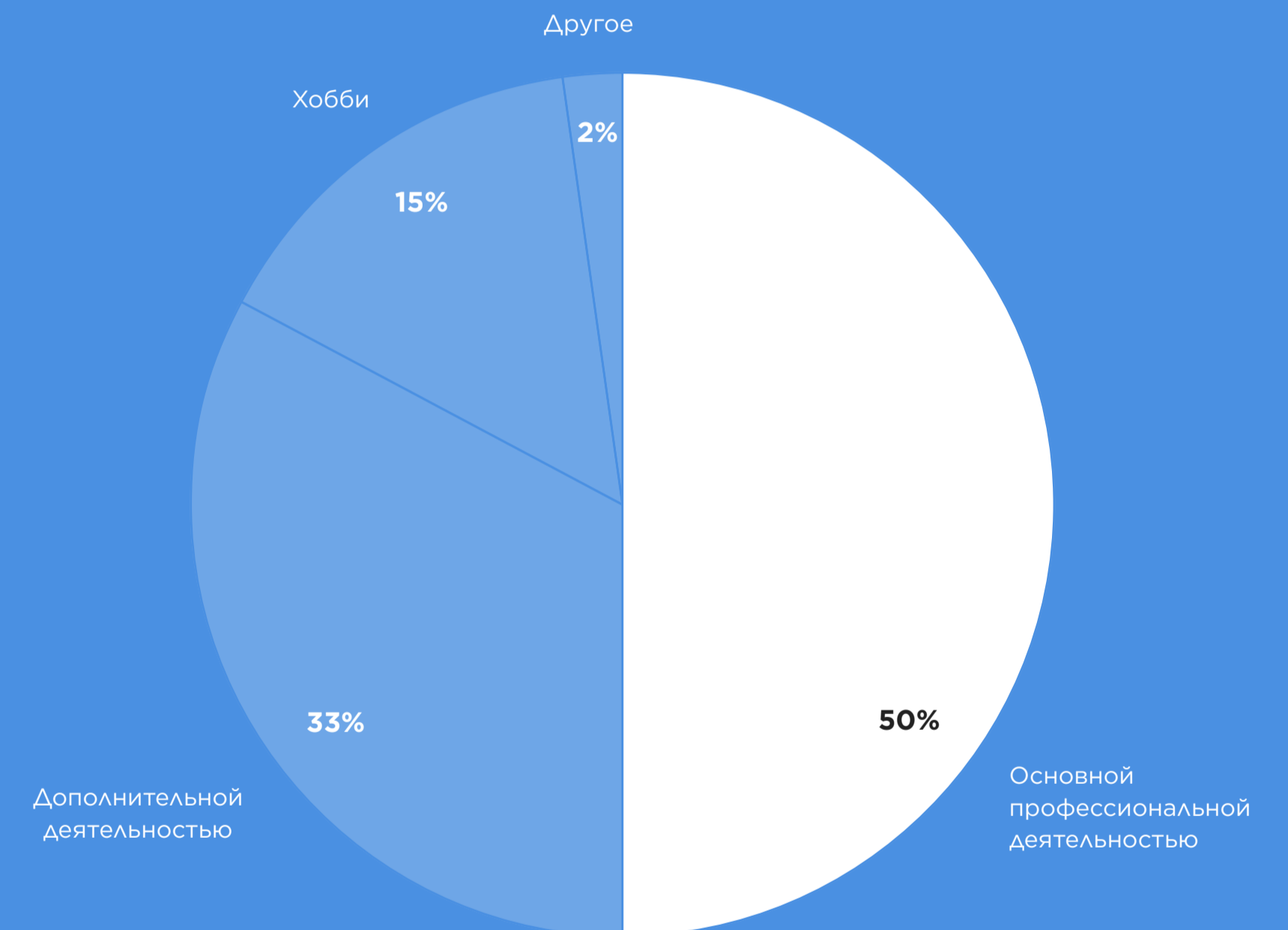
Анализ данных — это вообще профессия?

Многие люди решают задачи анализа данных наряду с программированием и другими должностными обязанностями. В России это особенно заметно: лишь 50% респондентов (в мире 65%, согласно Kaggle*) указали анализ данных как основную профессиональную деятельность, для 33% опрошенных — это дополнительная работа.

Доля Scala/Java-разработчиков среди тех, для кого анализ данных — это основная деятельность, почти в три раза меньше, чем среди тех, для кого дополнительная (18% против 46%). Это свидетельствует о том, что многие разработчики по долгу службы вынуждены решать задачи, связанные с обработкой данных.

Анализ данных в основном является для вас...

Все Python R Scala Java



1. Профиль специалиста в области Data Science

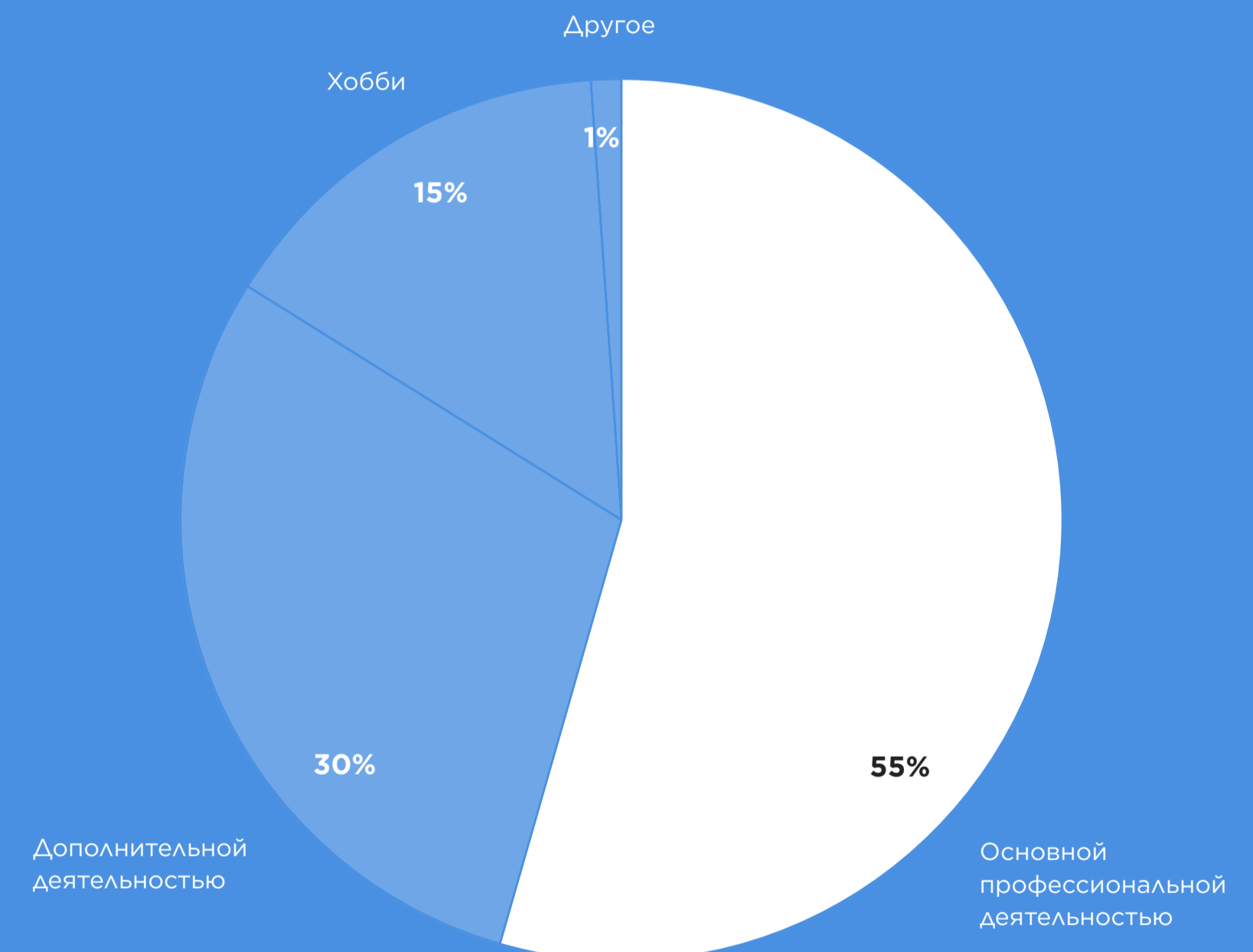
Анализ данных — это вообще профессия?

Многие люди решают задачи анализа данных наряду с программированием и другими должностными обязанностями. В России это особенно заметно: лишь 50% респондентов (в мире 65%, согласно Kaggle*) указали анализ данных как основную профессиональную деятельность, для 33% опрошенных — это дополнительная работа.

Доля Scala/Java-разработчиков среди тех, для кого анализ данных — это основная деятельность, почти в три раза меньше, чем среди тех, для кого дополнительная (18% против 46%). Это свидетельствует о том, что многие разработчики по долгу службы вынуждены решать задачи, связанные с обработкой данных.

Анализ данных в основном является для вас...

Все Python R Scala Java



1. Профиль специалиста в области Data Science

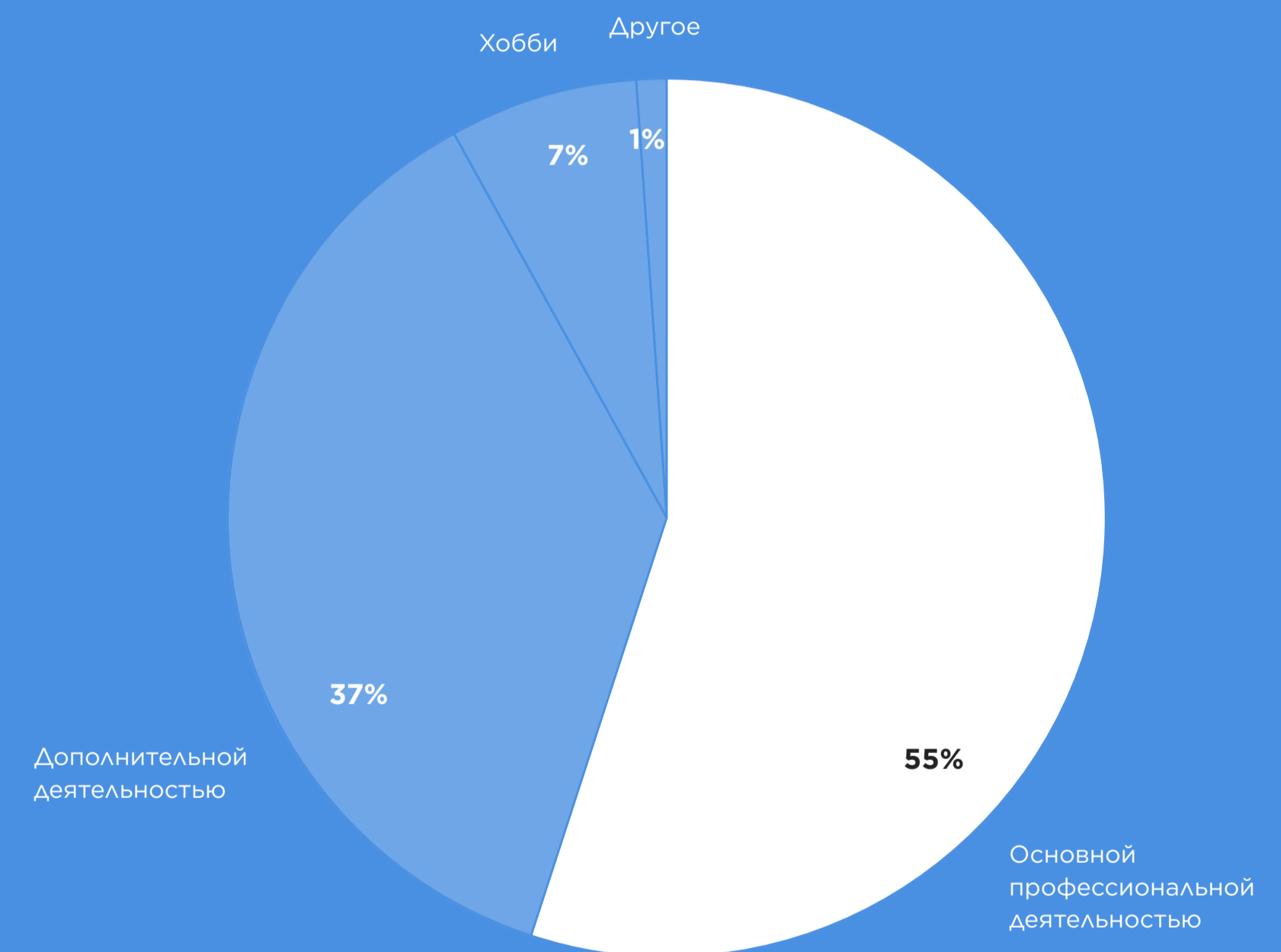
Анализ данных — это вообще профессия?

Многие люди решают задачи анализа данных наряду с программированием и другими должностными обязанностями. В России это особенно заметно: лишь 50% респондентов (в мире 65%, согласно Kaggle*) указали анализ данных как основную профессиональную деятельность, для 33% опрошенных — это дополнительная работа.

Доля Scala/Java-разработчиков среди тех, для кого анализ данных — это основная деятельность, почти в три раза меньше, чем среди тех, для кого дополнительная (18% против 46%). Это свидетельствует о том, что многие разработчики по долгу службы вынуждены решать задачи, связанные с обработкой данных.

Анализ данных в основном является для вас...

Все Python R Scala Java



1. Профиль специалиста в области Data Science

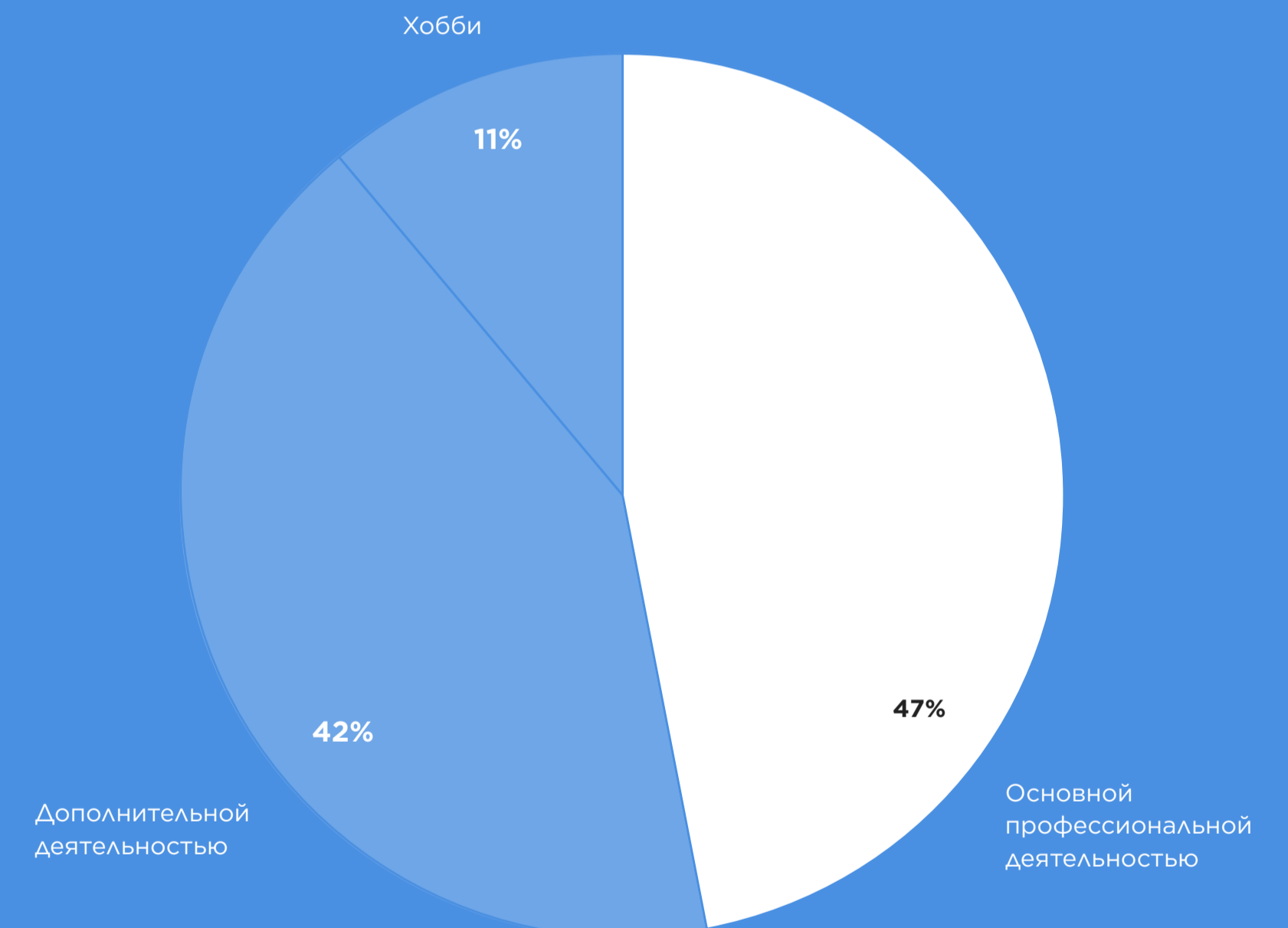
Анализ данных — это вообще профессия?

Многие люди решают задачи анализа данных наряду с программированием и другими должностными обязанностями. В России это особенно заметно: лишь 50% респондентов (в мире 65%, согласно Kaggle*) указали анализ данных как основную профессиональную деятельность, для 33% опрошенных — это дополнительная работа.

Доля Scala/Java-разработчиков среди тех, для кого анализ данных — это основная деятельность, почти в три раза меньше, чем среди тех, для кого дополнительная (18% против 46%). Это свидетельствует о том, что многие разработчики по долгу службы вынуждены решать задачи, связанные с обработкой данных.

Анализ данных в основном является для вас...

Все Python R Scala Java



1. Профиль специалиста в области Data Science

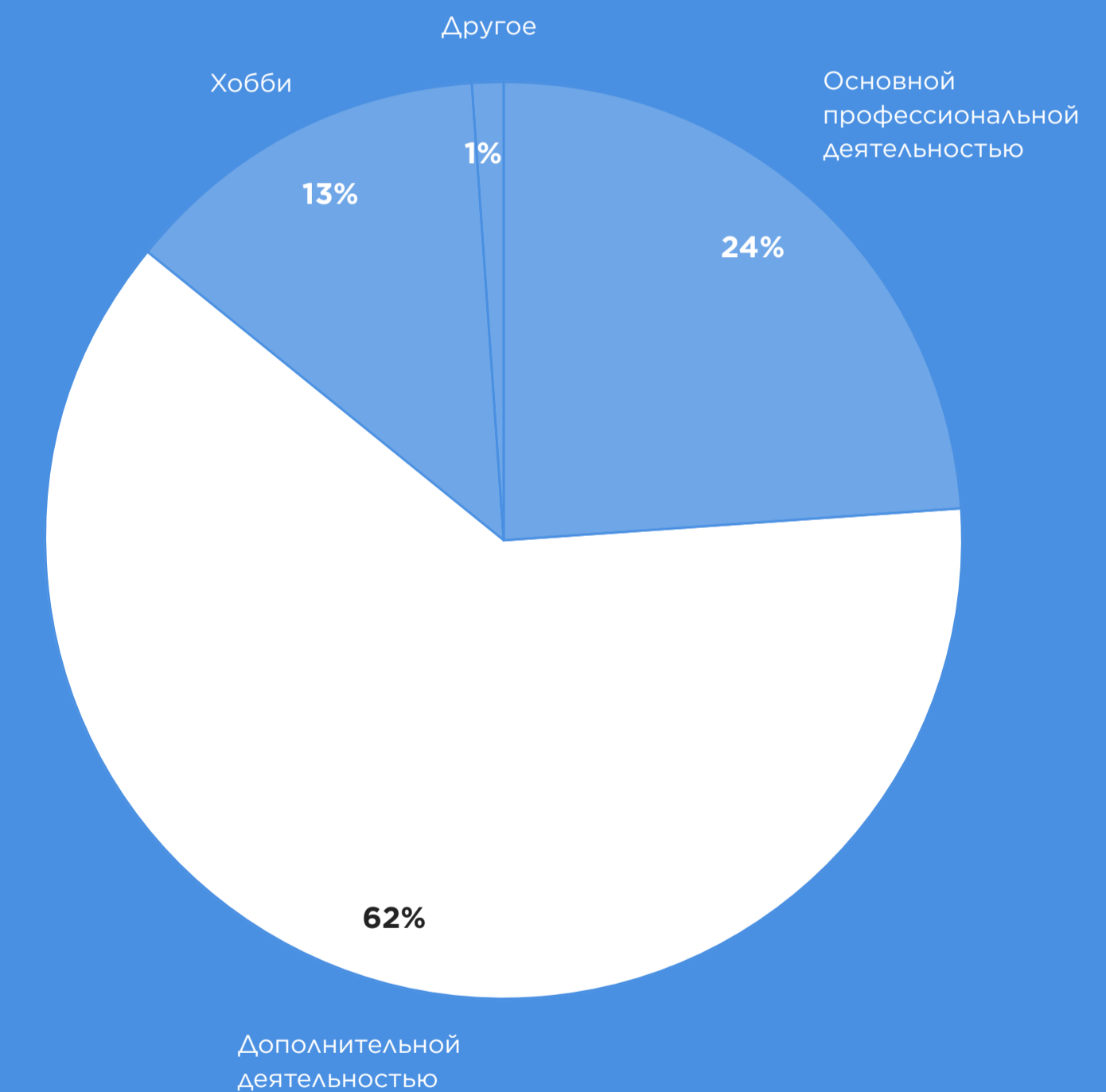
Анализ данных — это вообще профессия?

Многие люди решают задачи анализа данных наряду с программированием и другими должностными обязанностями. В России это особенно заметно: лишь 50% респондентов (в мире 65%, согласно Kaggle*) указали анализ данных как основную профессиональную деятельность, для 33% опрошенных — это дополнительная работа.

Доля Scala/Java-разработчиков среди тех, для кого анализ данных — это основная деятельность, почти в три раза меньше, чем среди тех, для кого дополнительная (18% против 46%). Это свидетельствует о том, что многие разработчики по долгу службы вынуждены решать задачи, связанные с обработкой данных.

Анализ данных в основном является для вас...

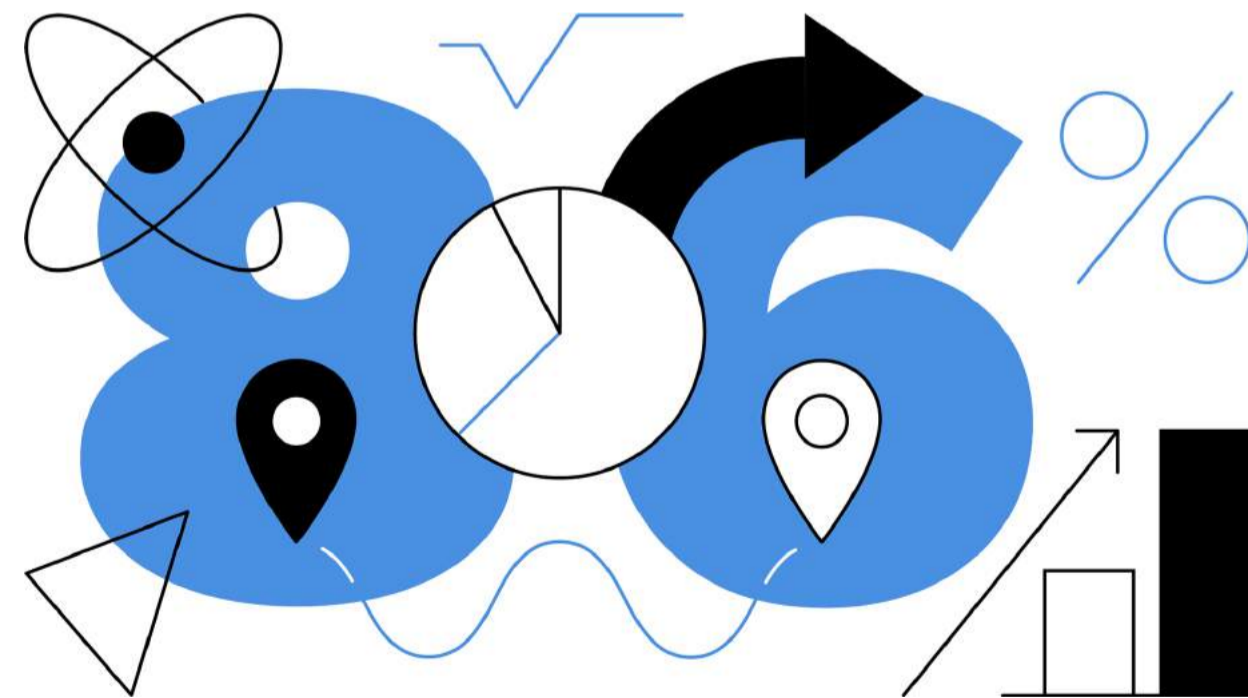
Все Python R Scala Java



1. Профиль специалиста в области Data Science

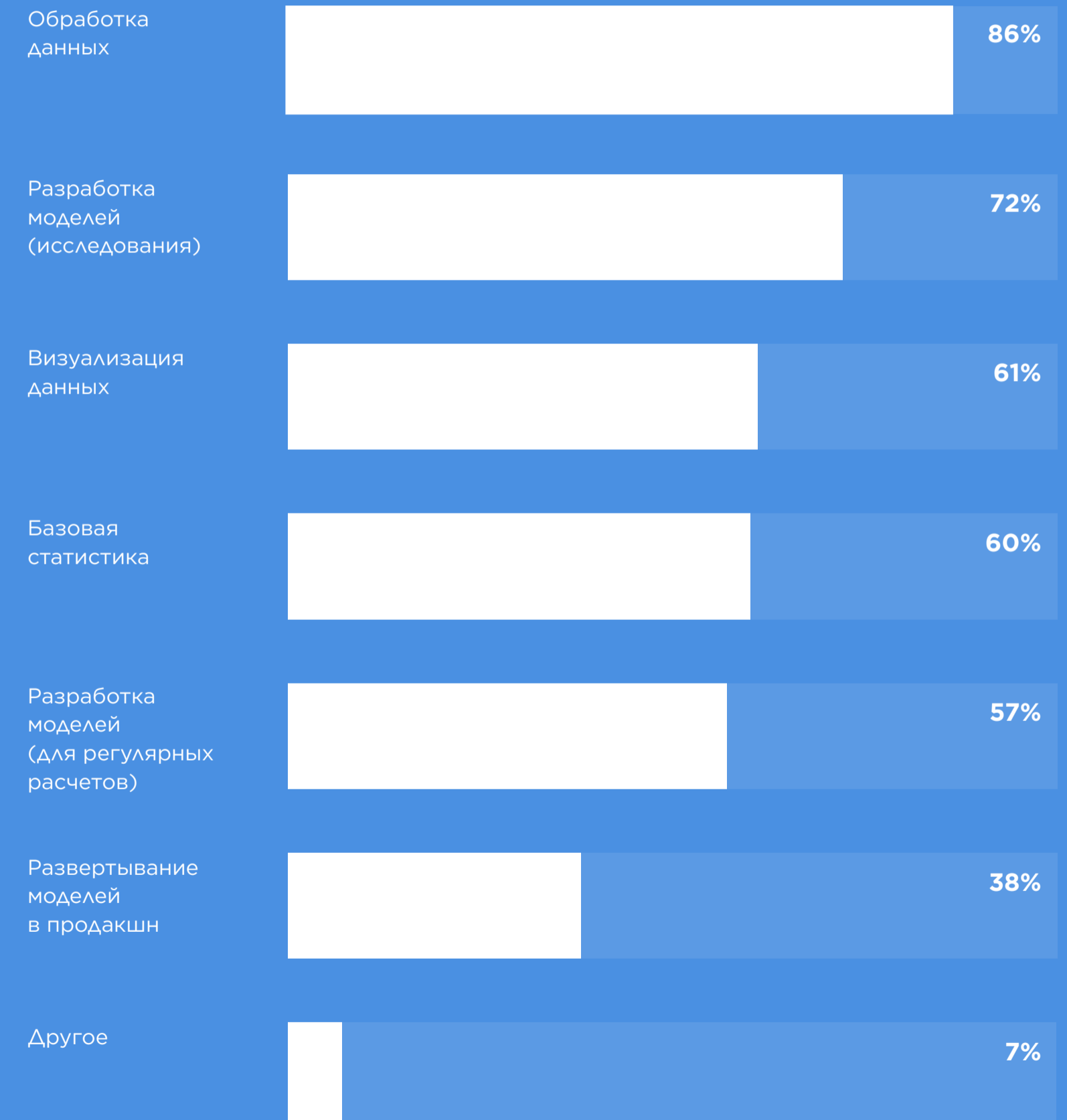
Анализ данных — это вообще профессия?

Анализ данных в России имеет свою специфику. Хотя обычно выделяют до 10 различных функций (job roles), вопрос разделения труда в анализе данных все еще довольно сложен.



опрошенных занимаются обработкой данных

В круг обязанностей или интересов входит



2. Языки программирования

Python — доминирующий язык анализа данных в России

Python — основной язык программирования у 74% респондентов. R в качестве основного языка указали только 10%. Можно ожидать, что в ближайшие годы будет расти, т. к. он является основным языком программирования для большинства опрошенных нами студентов (90% против 5% у R).

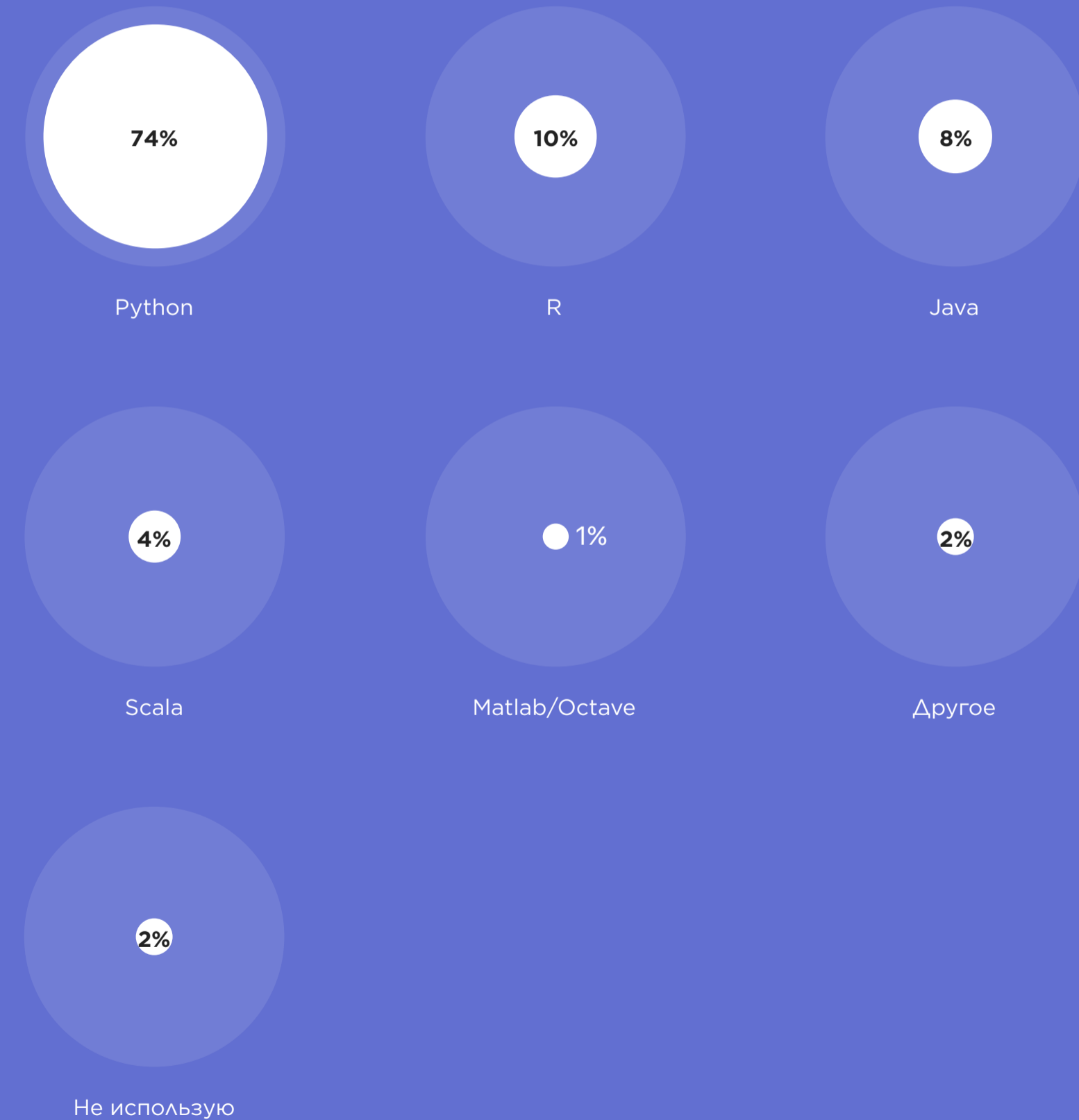


Дмитрий Бугайченко
инженер-аналитик, OK.RU,
Санкт-Петербург

«Конечно, доминирование питона несколько расстраивает, но против течения идти бесполезно. Весьма вероятно, что в итоге схема со строгим разделением обязанностей на разработку и внедрение в долгосрочной перспективе будет все больше доминировать».

Основной язык программирования

Все Big Data No Big Data Студенты



2. Языки программирования

Python — доминирующий язык анализа данных в России

Python — основной язык программирования у 74% респондентов. R в качестве основного языка указали только 10%. Можно ожидать, что в ближайшие годы будет расти, т. к. он является основным языком программирования для большинства опрошенных нами студентов (90% против 5% у R).



Дмитрий Бугайченко
инженер-аналитик, OK.RU,
Санкт-Петербург

«Конечно, доминирование питона несколько расстраивает, но против течения идти бесполезно. Весьма вероятно, что в итоге схема со строгим разделением обязанностей на разработку и внедрение в долгосрочной перспективе будет все больше доминировать».

Основной язык программирования

Все Big Data No Big Data Студенты



2. Языки программирования

Python — доминирующий язык анализа данных в России

Python — основной язык программирования у 74% респондентов. R в качестве основного языка указали только 10%. Можно ожидать, что в ближайшие годы будет расти, т. к. он является основным языком программирования для большинства опрошенных нами студентов (90% против 5% у R).

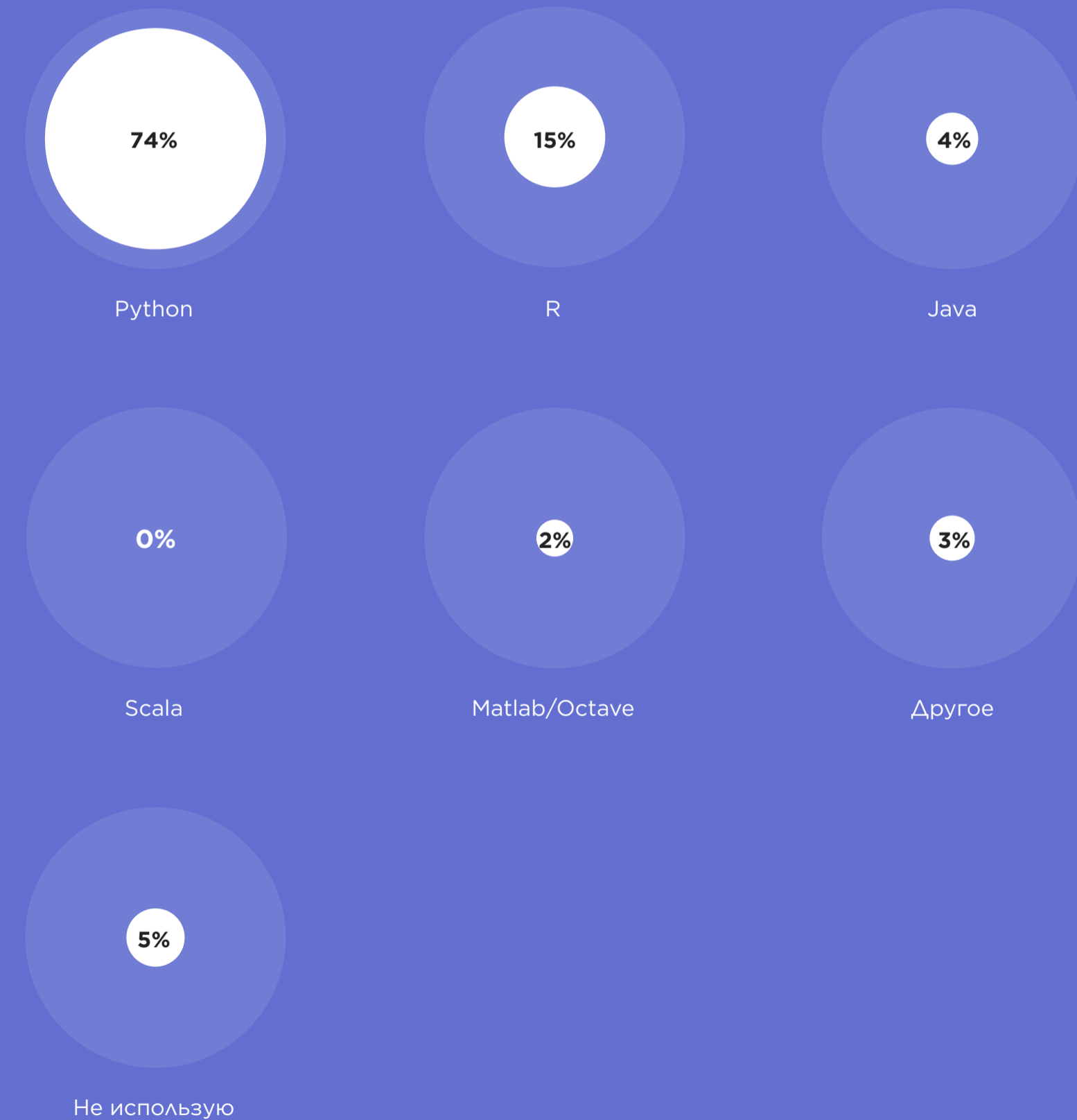


Дмитрий Бугайченко
инженер-аналитик, OK.RU,
Санкт-Петербург

«Конечно, доминирование питона несколько расстраивает, но против течения идти бесполезно. Весьма вероятно, что в итоге схема со строгим разделением обязанностей на разработку и внедрение в долгосрочной перспективе будет все больше доминировать».

Основной язык программирования

Все Big Data No Big Data Студенты



2. Языки программирования

Python — доминирующий язык анализа данных в России

Python — основной язык программирования у 74% респондентов. R в качестве основного языка указали только 10%. Можно ожидать, что в ближайшие годы будет расти, т. к. он является основным языком программирования для большинства опрошенных нами студентов (90% против 5% у R).

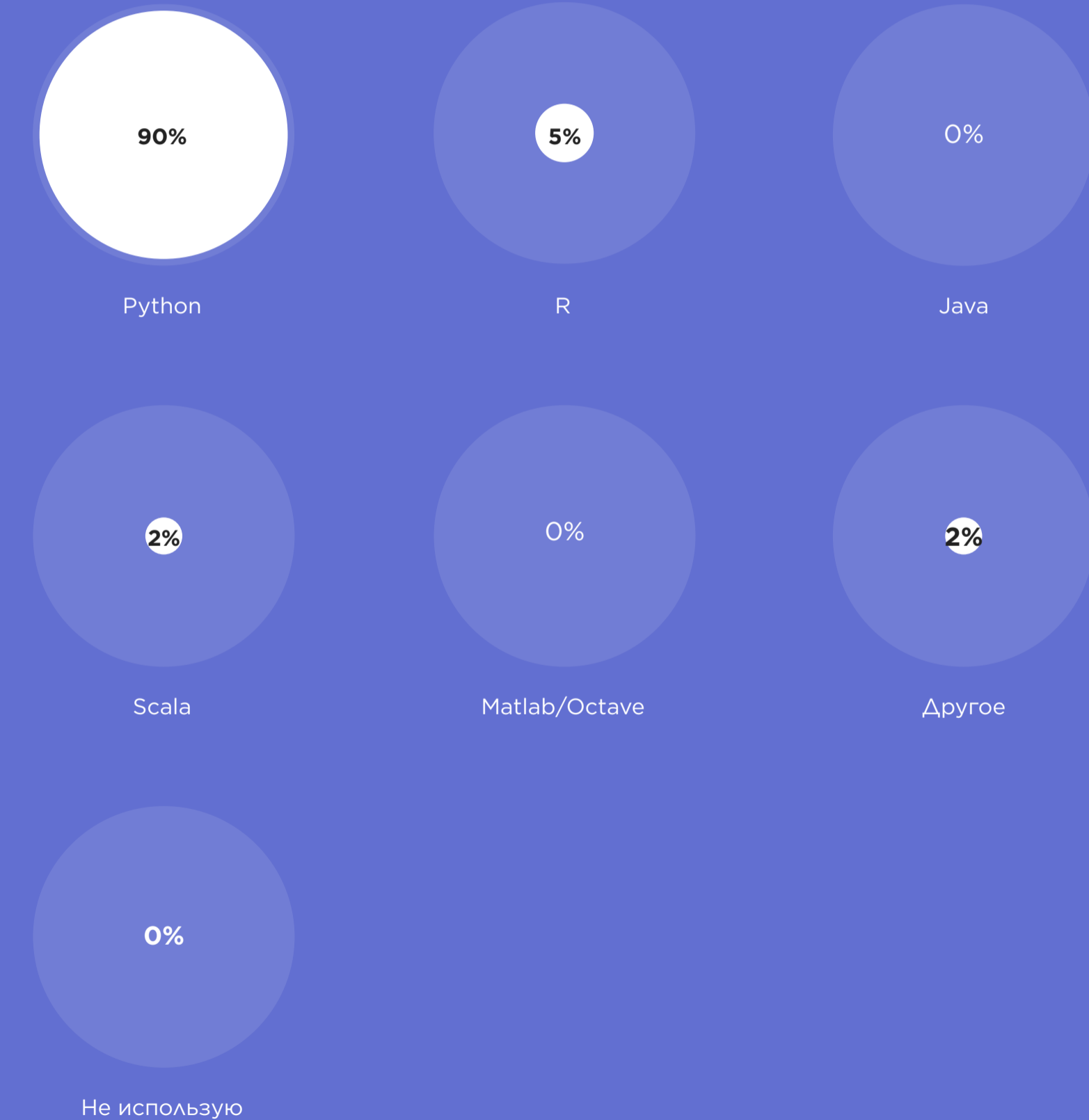


Дмитрий Бугайченко
инженер-аналитик, OK.RU,
Санкт-Петербург

«Конечно, доминирование питона несколько расстраивает, но против течения идти бесполезно. Весьма вероятно, что в итоге схема со строгим разделением обязанностей на разработку и внедрение в долгосрочной перспективе будет все больше доминировать».

Основной язык программирования

Все Big Data No Big Data Студенты



2. Языки программирования

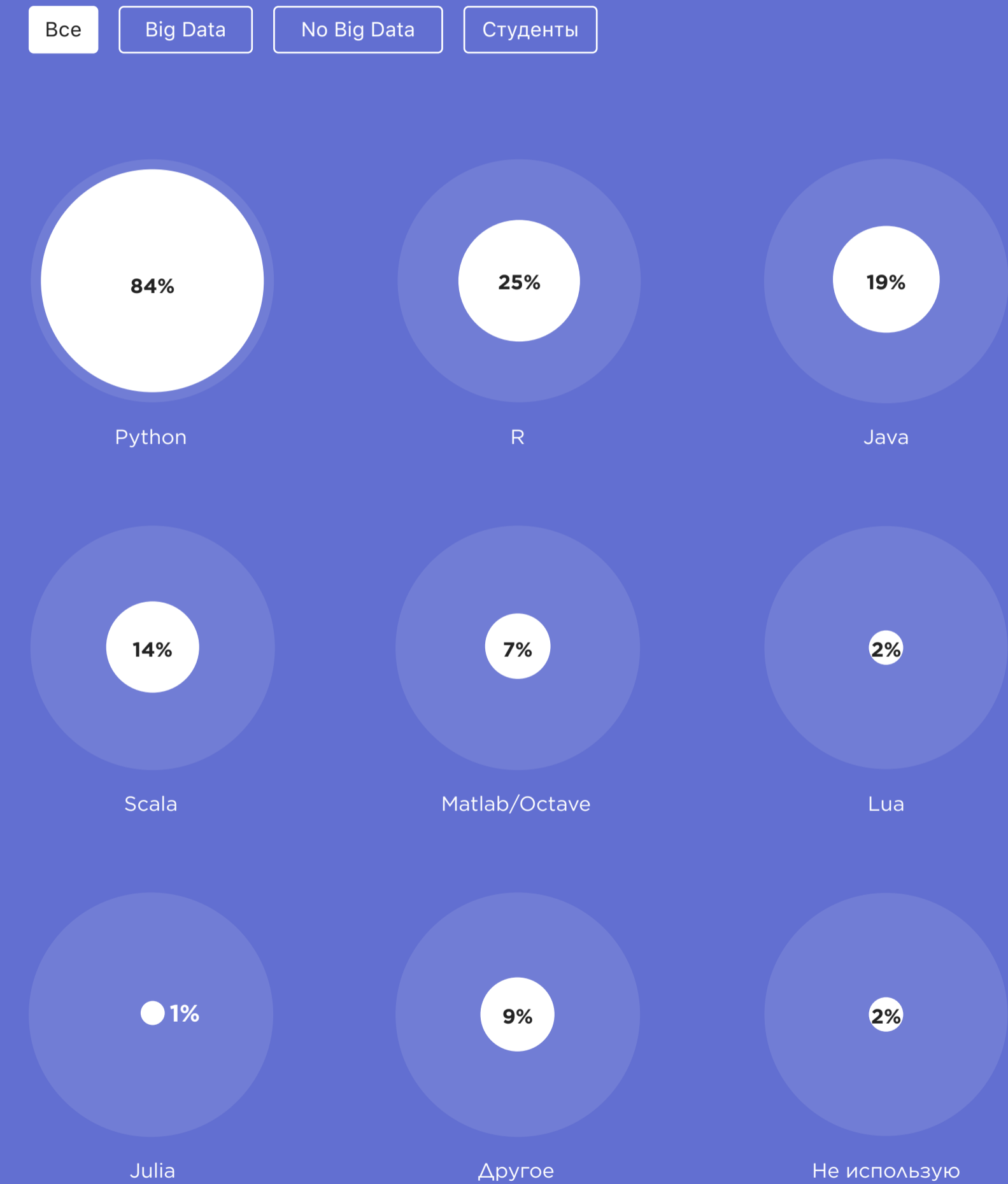
Python — доминирующий язык анализа данных в России

В отличие от ситуации в мире, где доли использования Python и R в области анализа данных примерно равны, в России Python значительно популярнее R — 84% против 25%.

В этой таблице мы разбили всех респондентов на категории по тому, как они ответили на вопрос об использовании инструментов для Big Data. Всех, кто использует подобные инструменты, мы отнесли к Big Data, а остальных — к No BigData. Студенты выделены в самостоятельную группу.

Самая большая доля JVM-языков — именно среди занимающихся «большими данными».

Регулярно используемые языки



2. Языки программирования

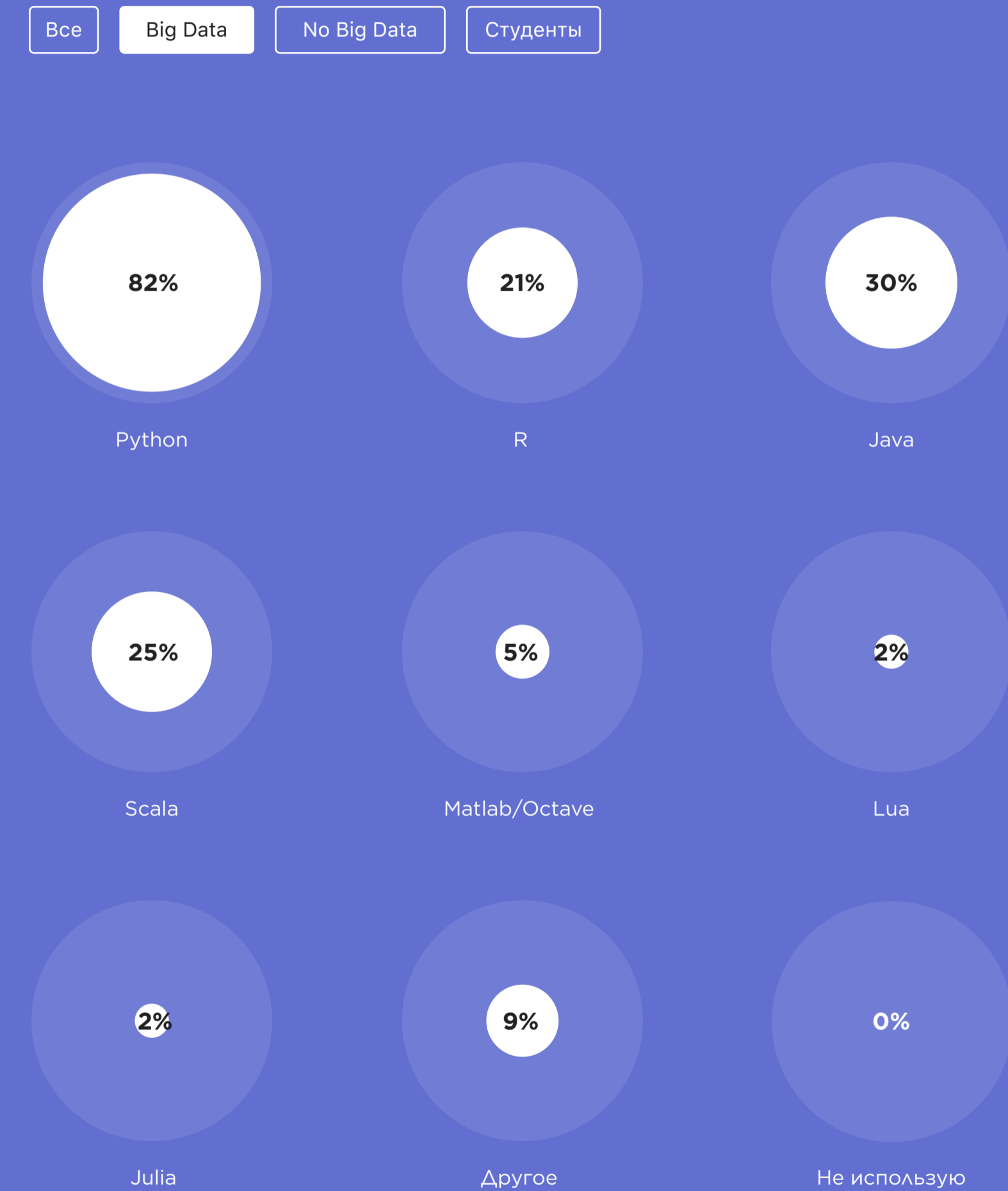
Python — доминирующий язык анализа данных в России

В отличие от ситуации в мире, где доли использования Python и R в области анализа данных примерно равны, в России Python значительно популярнее R — 84% против 25%.

В этой таблице мы разбили всех респондентов на категории по тому, как они ответили на вопрос об использовании инструментов для Big Data. Всех, кто использует подобные инструменты, мы отнесли к Big Data, а остальных — к No BigData. Студенты выделены в самостоятельную группу.

Самая большая доля JVM-языков — именно среди занимающихся «большими данными».

Регулярно используемые языки



2. Языки программирования

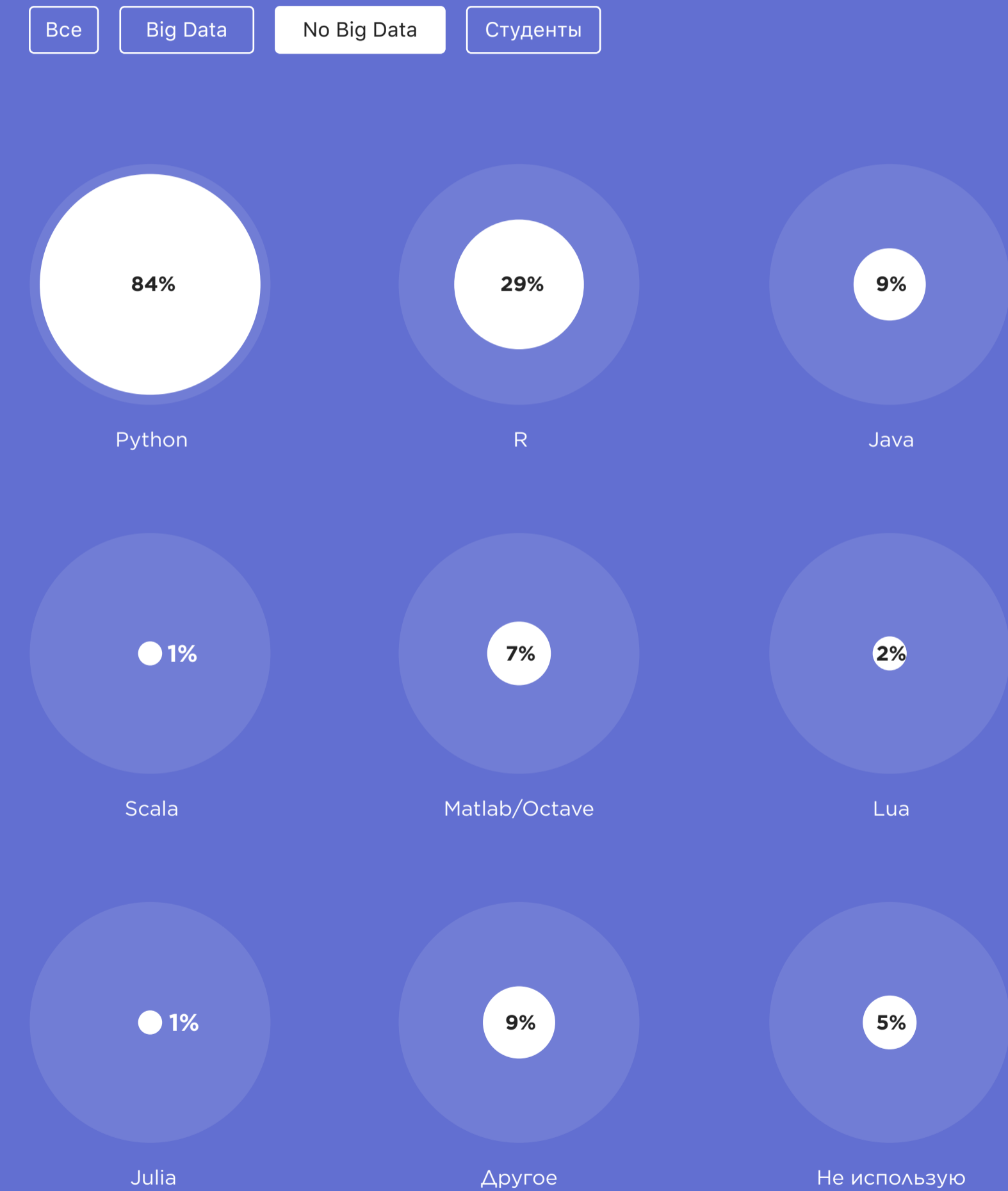
Python — доминирующий язык анализа данных в России

В отличие от ситуации в мире, где доли использования Python и R в области анализа данных примерно равны, в России Python значительно популярнее R — 84% против 25%.

В этой таблице мы разбили всех респондентов на категории по тому, как они ответили на вопрос об использовании инструментов для Big Data. Всех, кто использует подобные инструменты, мы отнесли к Big Data, а остальных — к No BigData. Студенты выделены в самостоятельную группу.

Самая большая доля JVM-языков — именно среди занимающихся «большими данными».

Регулярно используемые языки



2. Языки программирования

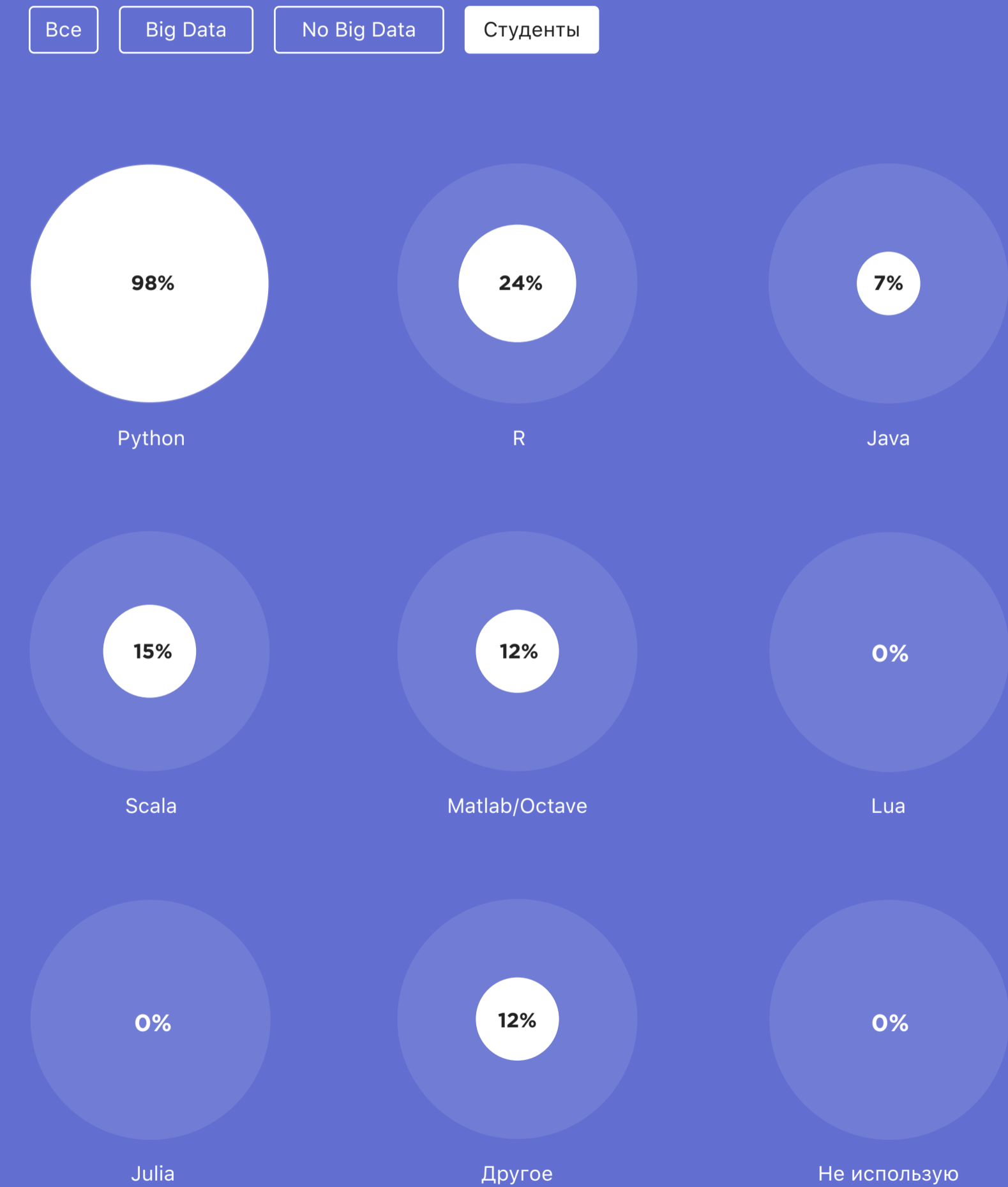
Python — доминирующий язык анализа данных в России

В отличие от ситуации в мире, где доли использования Python и R в области анализа данных примерно равны, в России Python значительно популярнее R — 84% против 25%.

В этой таблице мы разбили всех респондентов на категории по тому, как они ответили на вопрос об использовании инструментов для Big Data. Всех, кто использует подобные инструменты, мы отнесли к Big Data, а остальных — к No BigData. Студенты выделены в самостоятельную группу.

Самая большая доля JVM-языков — именно среди занимающихся «большими данными».

Регулярно используемые языки



2. Языки программирования

Больше половины тех, для кого R — основной язык программирования, также используют Python.

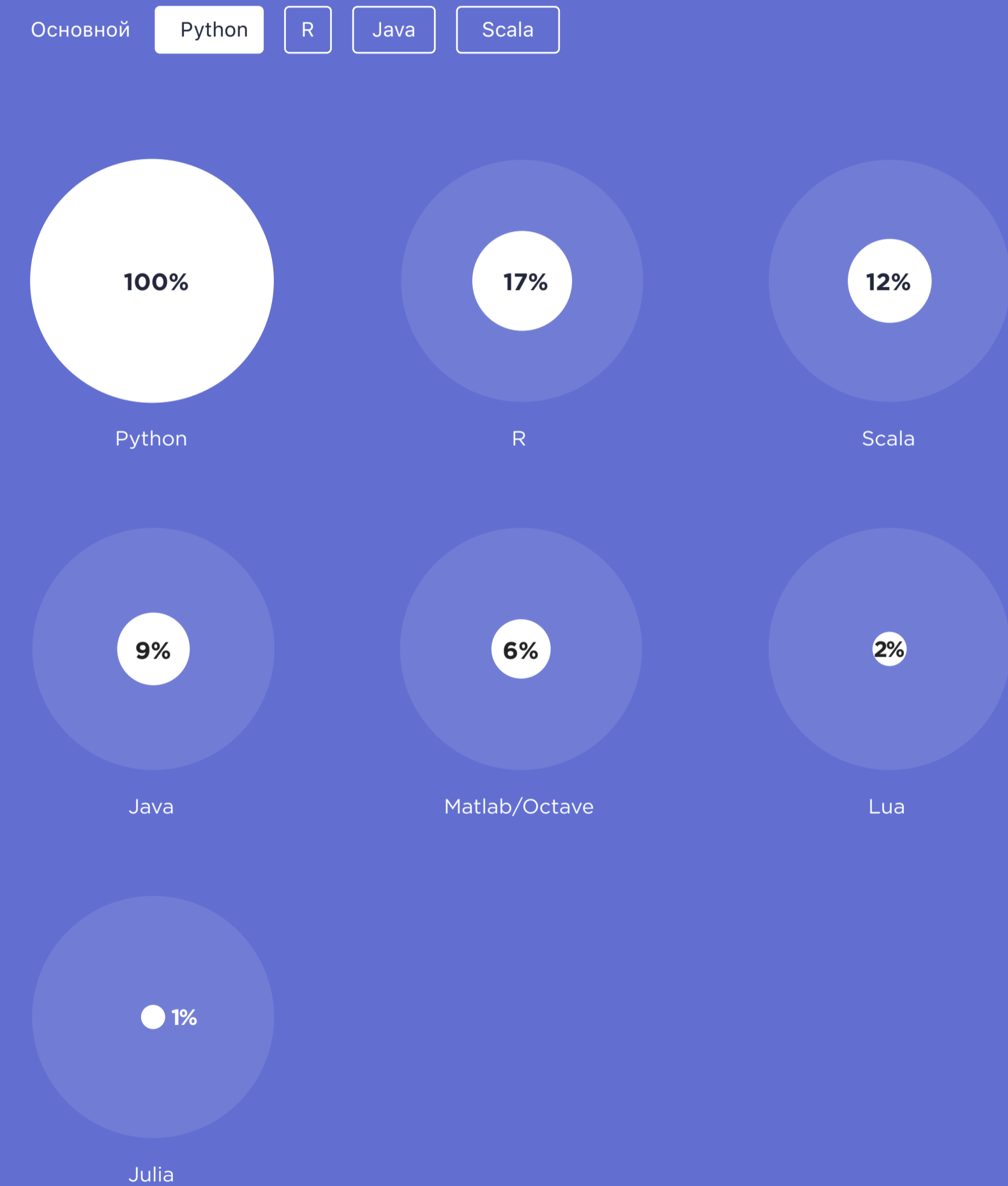
Респонденты-джависты используют мало дополнительных языков, в отличие от любителей Scala, которые активно пользуются и Java, и Python, и R. При этом нужно учесть, что тех, для кого Scala является основным языком, набралось мало.



Юрий Кашницкий
программист-исследователь
группы предиктивной аналитики,
Mail.ru Group, Москва

«Стоит отметить также, что Python — наиболее популярный язык в стремительно развивающихся библиотеках глубокого обучения (TensorFlow, Keras, PyTorch и т.д.), что тоже позволяет говорить о последующем укреплении позиций Python»

Пересечение для основных и регулярно используемых языков



2. Языки программирования

Больше половины тех, для кого R — основной язык программирования, также используют Python.

Респонденты-джависты используют мало дополнительных языков, в отличие от любителей Scala, которые активно пользуются и Java, и Python, и R. При этом нужно учесть, что тех, для кого Scala является основным языком, набралось мало.



Юрий Кашницкий
программист-исследователь
группы предиктивной аналитики,
Mail.ru Group, Москва

«Стоит отметить также, что Python — наиболее популярный язык в стремительно развивающихся библиотеках глубокого обучения (TensorFlow, Keras, PyTorch и т.д.), что тоже позволяет говорить о последующем укреплении позиций Python»

Пересечение для основных и регулярно используемых языков



2. Языки программирования

Больше половины тех, для кого R — основной язык программирования, также используют Python.

Респонденты-джависты используют мало дополнительных языков, в отличие от любителей Scala, которые активно пользуются и Java, и Python, и R. При этом нужно учесть, что тех, для кого Scala является основным языком, набралось мало.



Юрий Кашницкий
программист-исследователь
группы предиктивной аналитики,
Mail.ru Group, Москва

«Стоит отметить также, что Python — наиболее популярный язык в стремительно развивающихся библиотеках глубокого обучения (TensorFlow, Keras, PyTorch и т.д.), что тоже позволяет говорить о последующем укреплении позиций Python»

Пересечение для основных и регулярно используемых языков



2. Языки программирования

Больше половины тех, для кого R — основной язык программирования, также используют Python.

Респонденты-джависты используют мало дополнительных языков, в отличие от любителей Scala, которые активно пользуются и Java, и Python, и R. При этом нужно учесть, что тех, для кого Scala является основным языком, набралось мало.



Юрий Кашницкий
программист-исследователь
группы предиктивной аналитики,
Mail.ru Group, Москва

«Стоит отметить также, что Python — наиболее популярный язык в стремительно развивающихся библиотеках глубокого обучения (TensorFlow, Keras, PyTorch и т.д.), что тоже позволяет говорить о последующем укреплении позиций Python»

Пересечение для основных и регулярно используемых языков



3. Технологии и инструменты

Глубокое обучение — это тренд

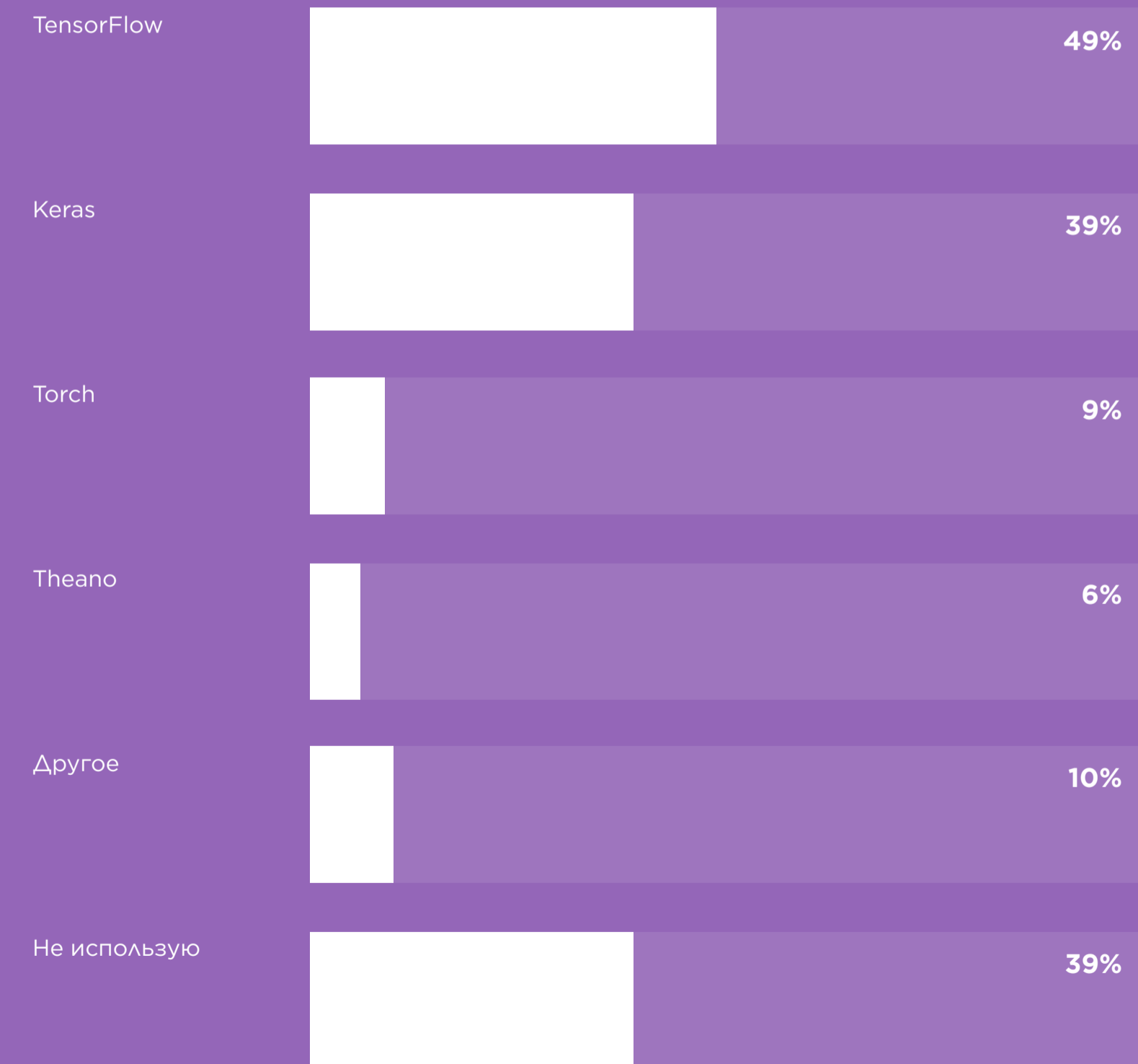
Более 60% респондентов так или иначе используют инструменты для глубокого обучения.

TensorFlow™ является наиболее популярным фреймворком — 49%, Keras на втором месте с 39%.

Глубокое обучение особенно популярно среди студентов — 83% опрошенных в этой группе используют фреймворки для глубокого обучения.

Какие инструменты для глубокого машинного обучения (deep learning) вы используете?

Все Студенты Не студенты



3. Технологии и инструменты

Глубокое обучение — это тренд

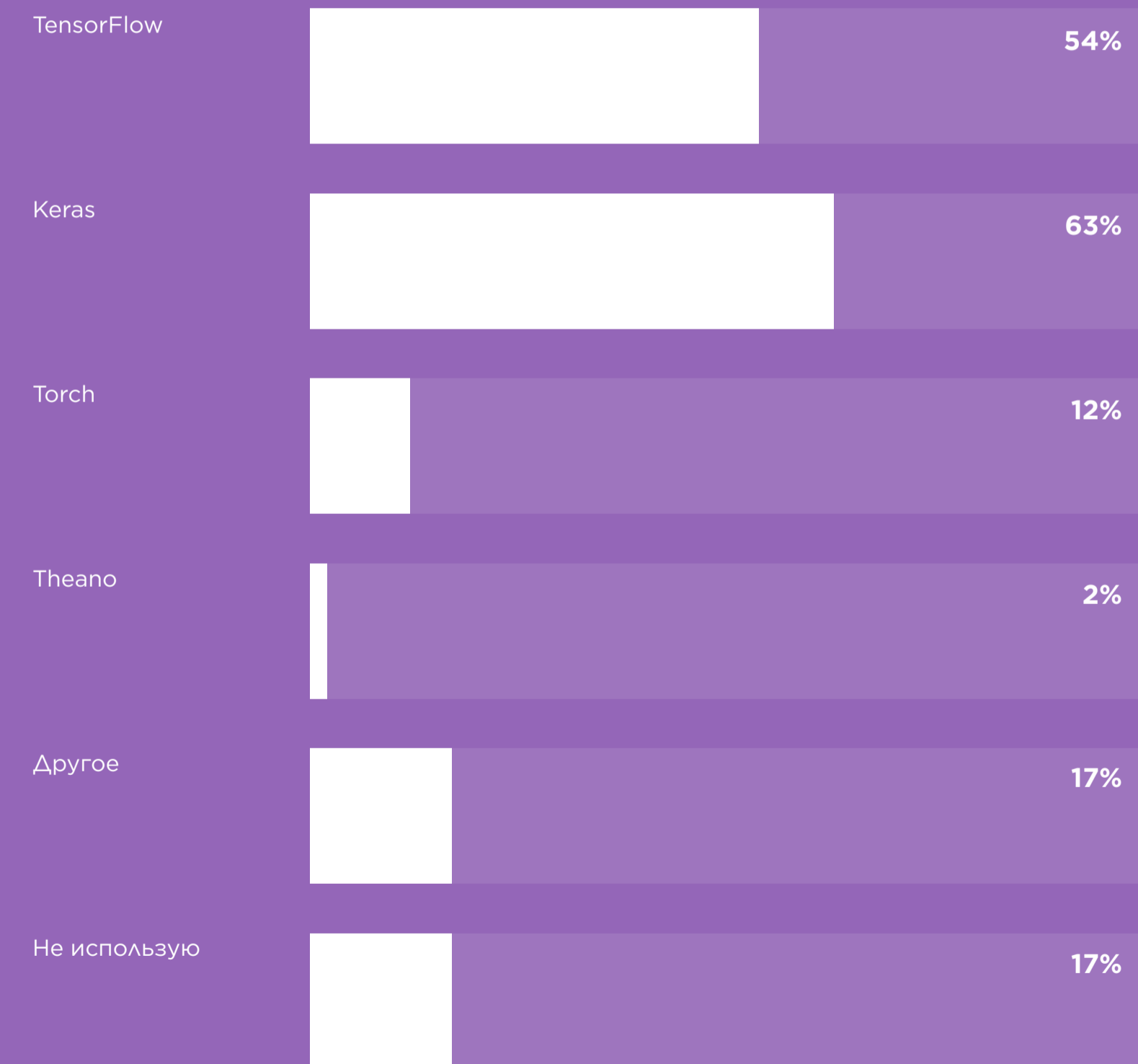
Более 60% респондентов так или иначе используют инструменты для глубокого обучения.

TensorFlow™ является наиболее популярным фреймворком — 49%, Keras на втором месте с 39%.

Глубокое обучение особенно популярно среди студентов — 83% опрошенных в этой группе используют фреймворки для глубокого обучения.

Какие инструменты для глубокого машинного обучения (deep learning) вы используете?

Все Студенты Не студенты



3. Технологии и инструменты

Глубокое обучение — это тренд

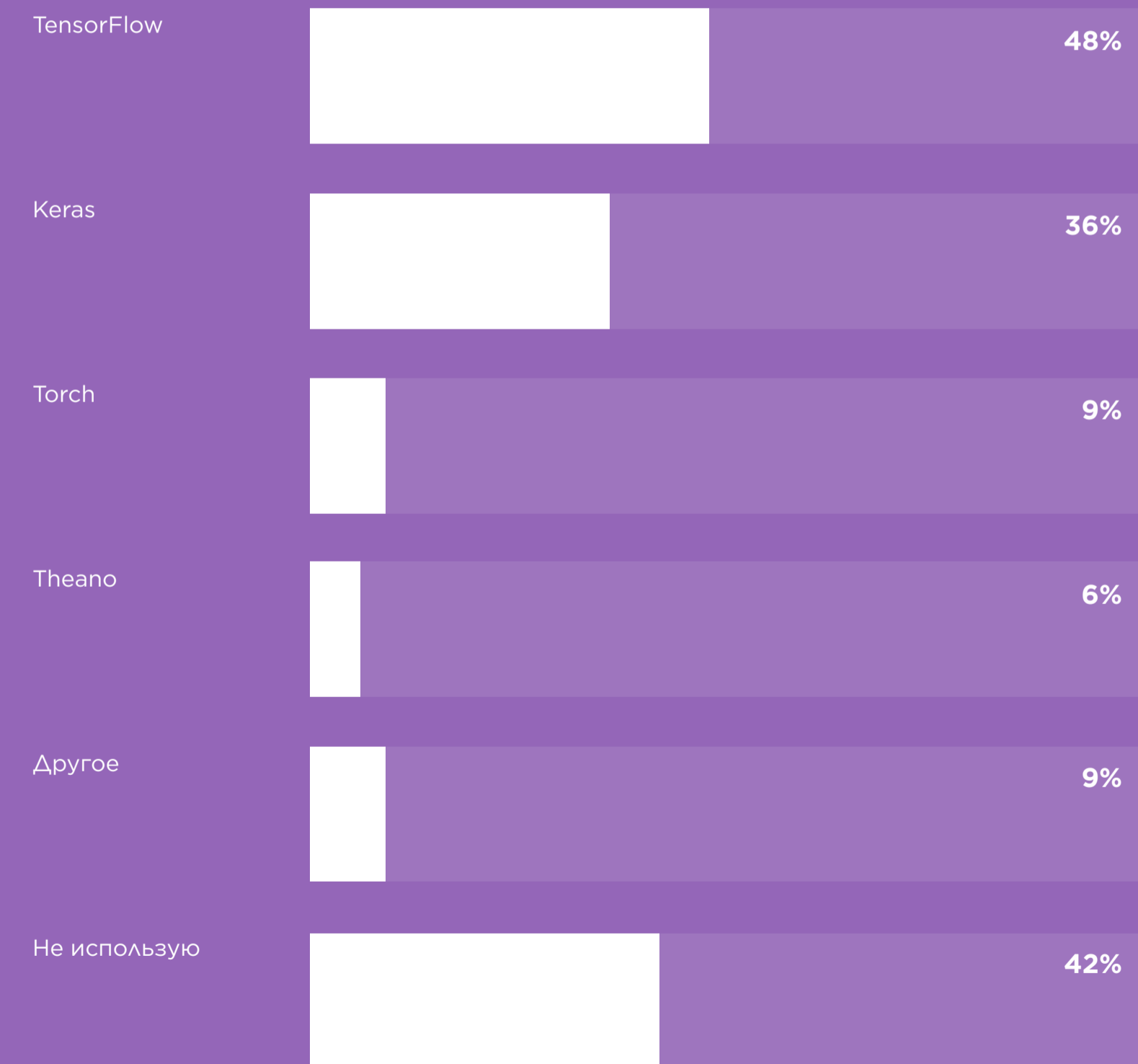
Более 60% респондентов так или иначе используют инструменты для глубокого обучения.

TensorFlow™ является наиболее популярным фреймворком — 49%, Keras на втором месте с 39%.

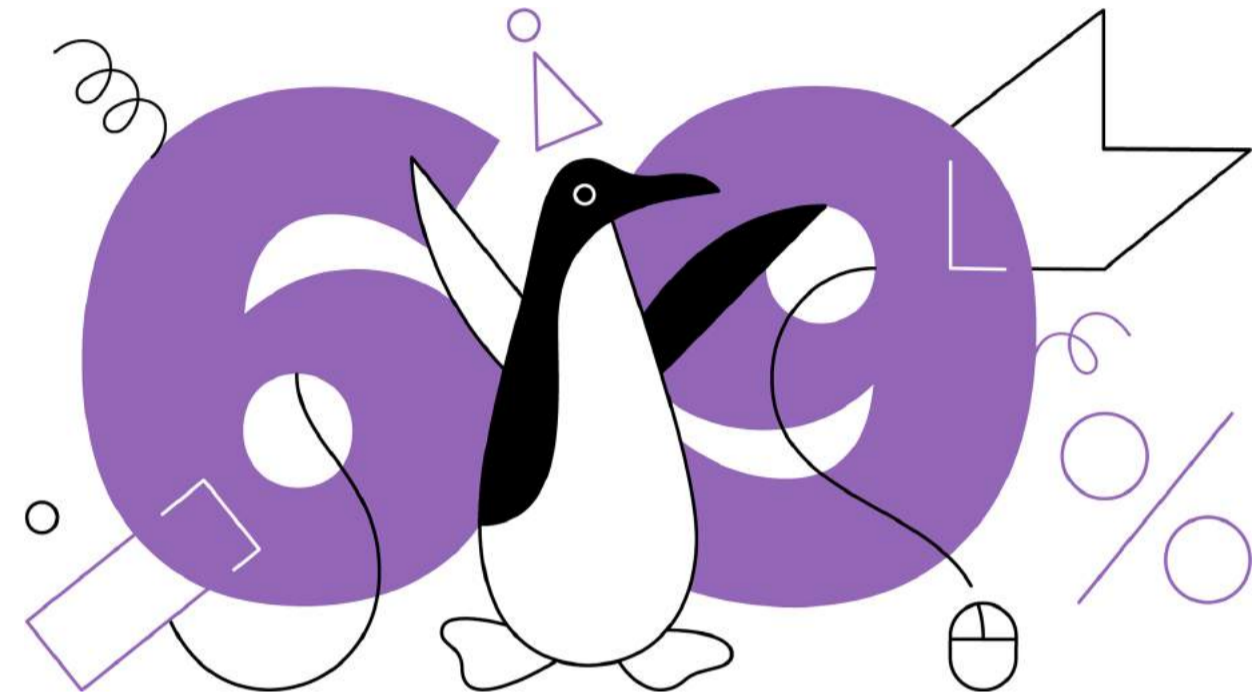
Глубокое обучение особенно популярно среди студентов — 83% опрошенных в этой группе используют фреймворки для глубокого обучения.

Какие инструменты для глубокого машинного обучения (deep learning) вы используете?

Все Студенты Не студенты



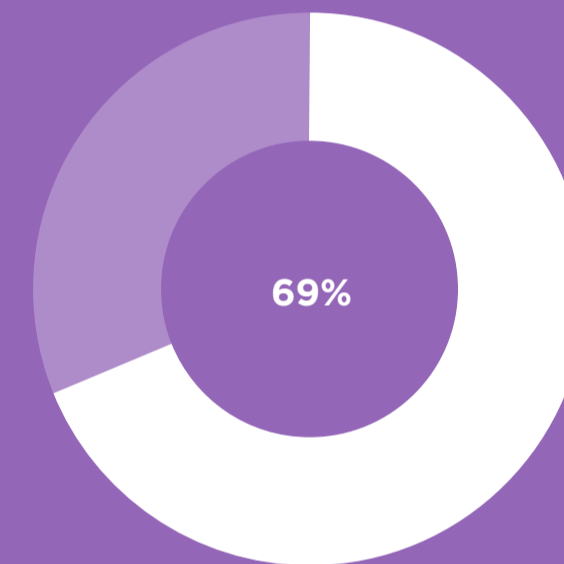
3. Технологии и инструменты



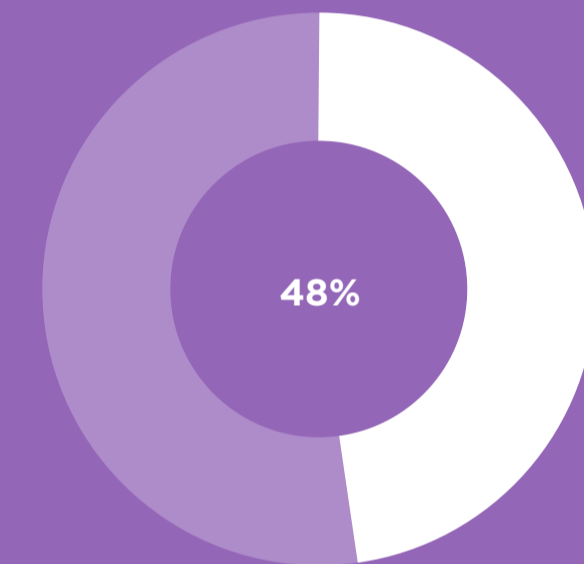
опрошенных используют
Linux для анализа данных

Какую операционную систему (системы) Вы используете в качестве рабочего окружения, когда занимаетесь анализом данных?

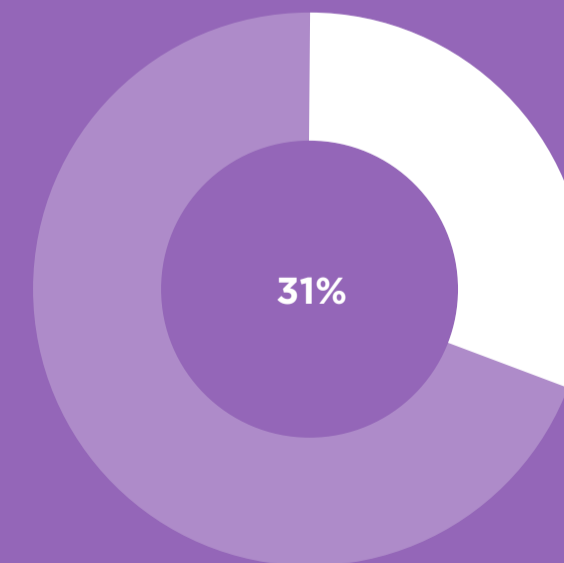
Все Студенты Не студенты



Linux

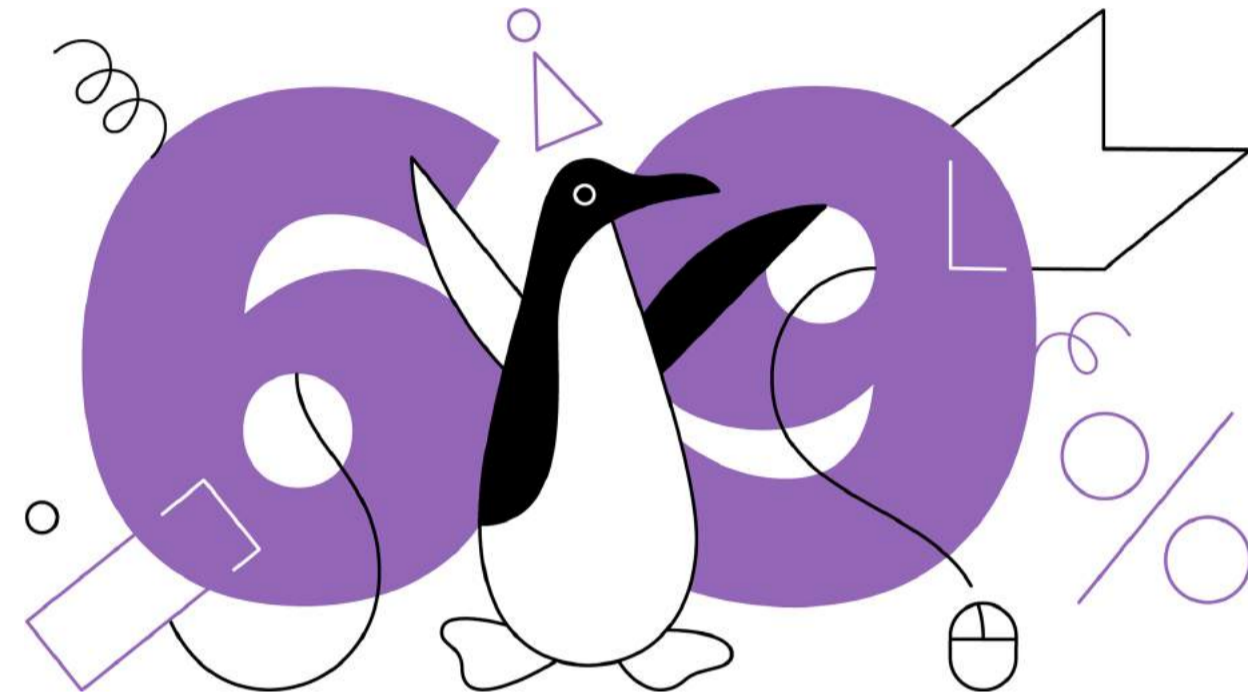


Windows



Mac OS X

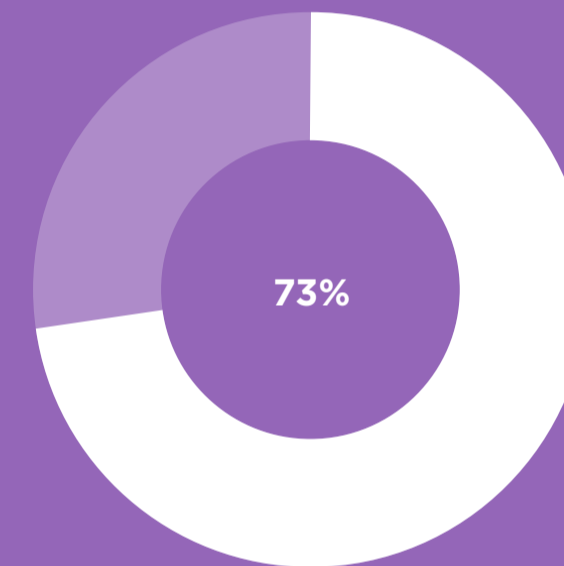
3. Технологии и инструменты



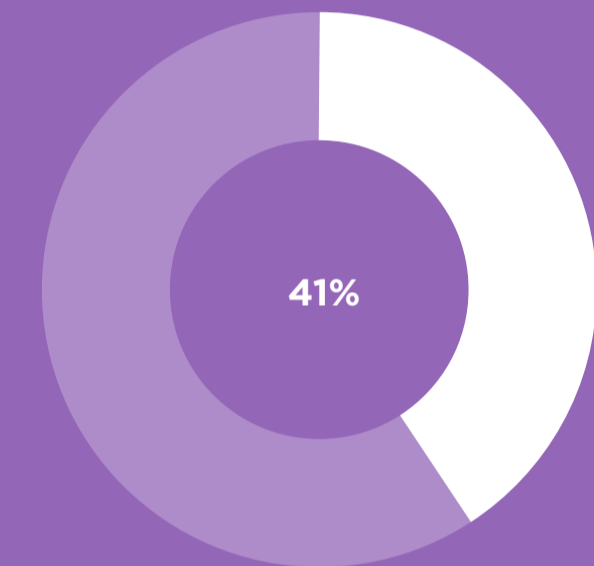
опрошенных используют
Linux для анализа данных

Какую операционную систему (системы) Вы используете в качестве рабочего окружения, когда занимаетесь анализом данных?

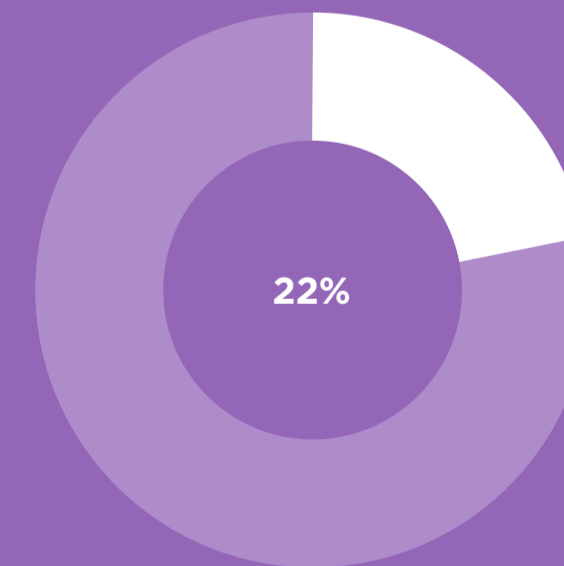
Все Студенты Не студенты



Linux

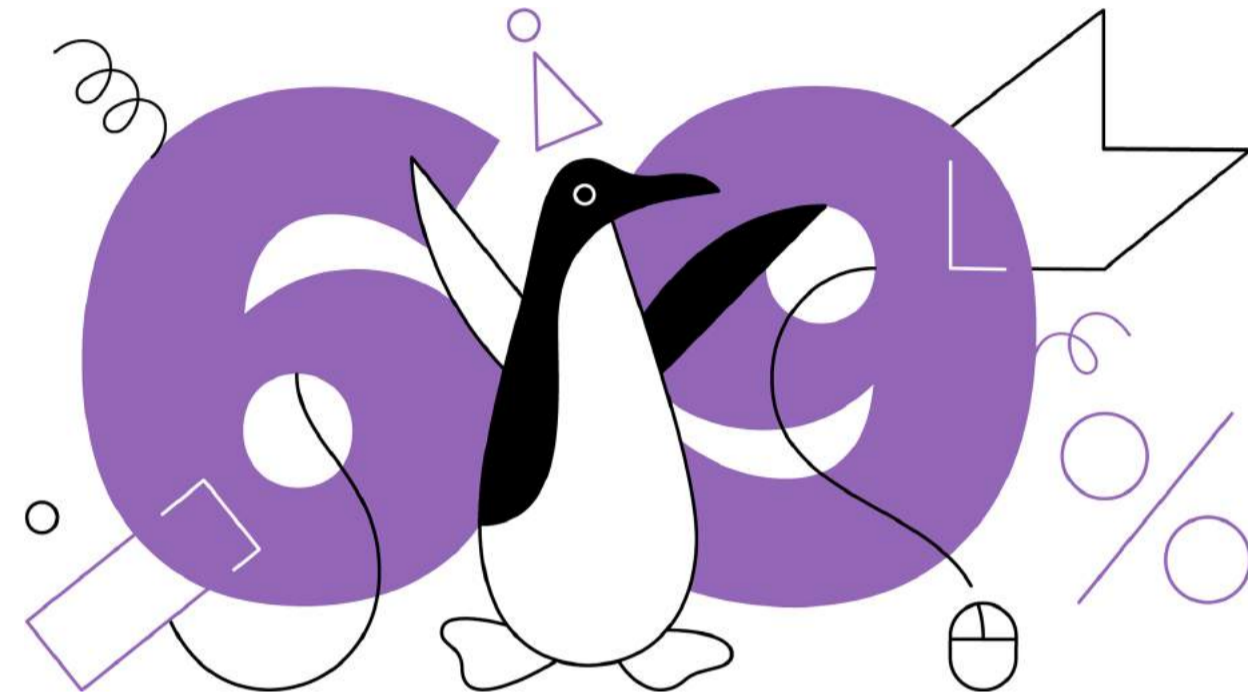


Windows



Mac OS X

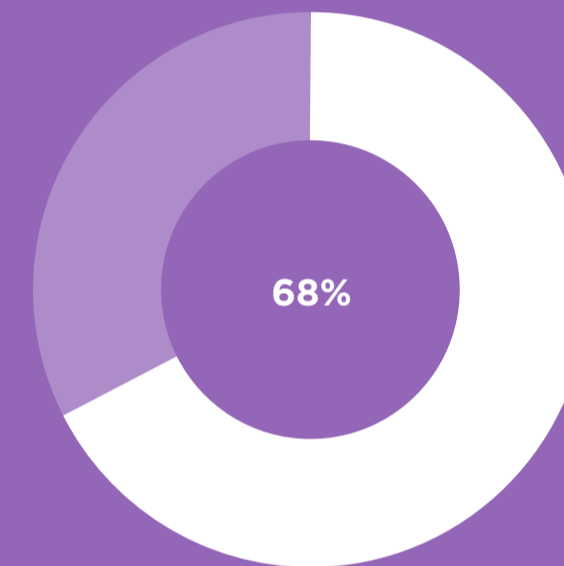
3. Технологии и инструменты



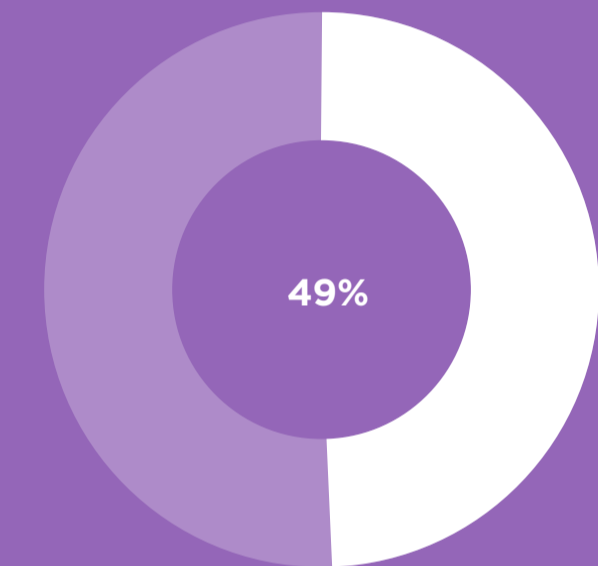
опрошенных используют
Linux для анализа данных

Какую операционную систему (системы) Вы используете в качестве рабочего окружения, когда занимаетесь анализом данных?

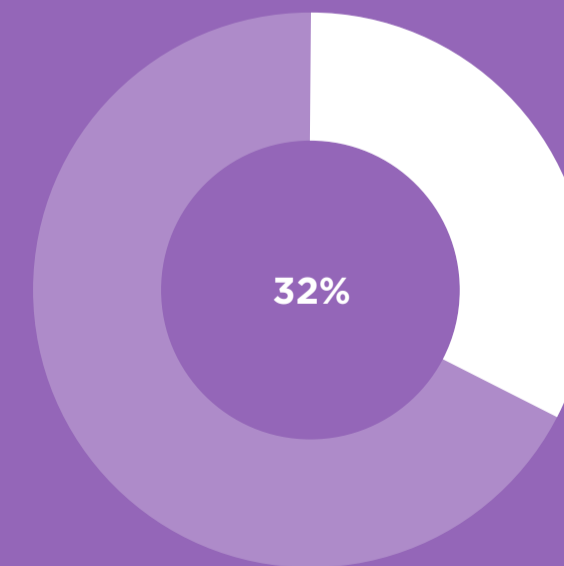
Все Студенты Не студенты



Linux

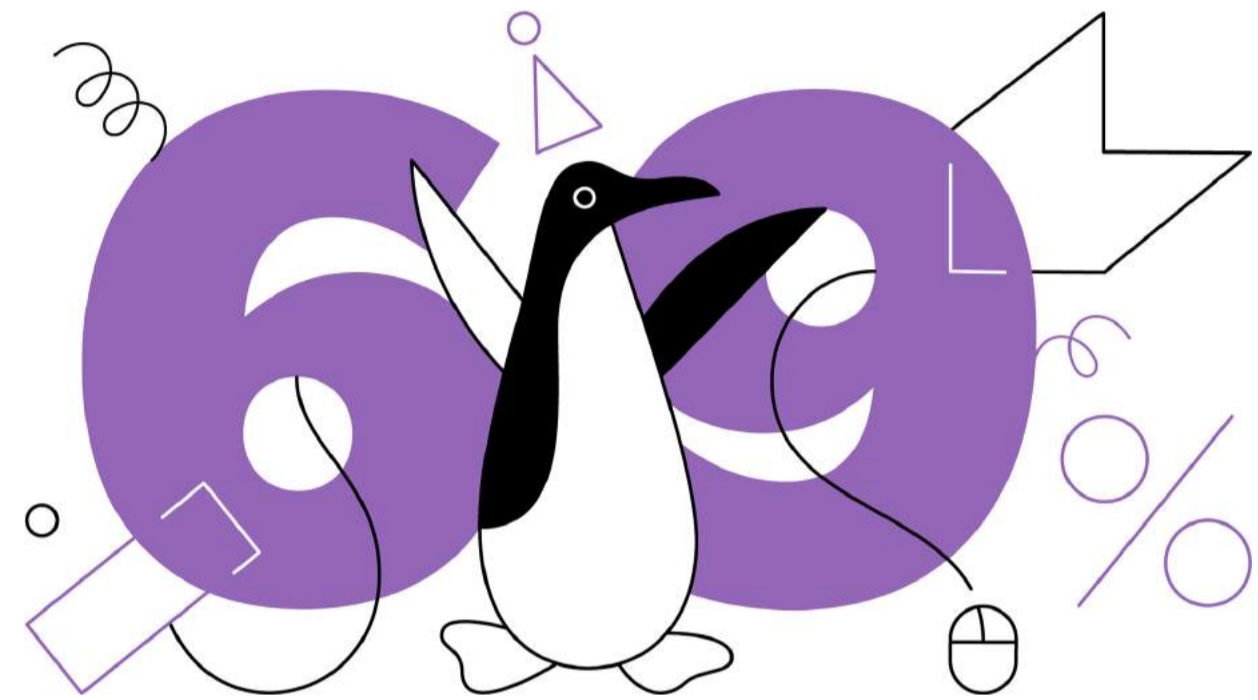


Windows



Mac OS X

3. Технологии и инструменты



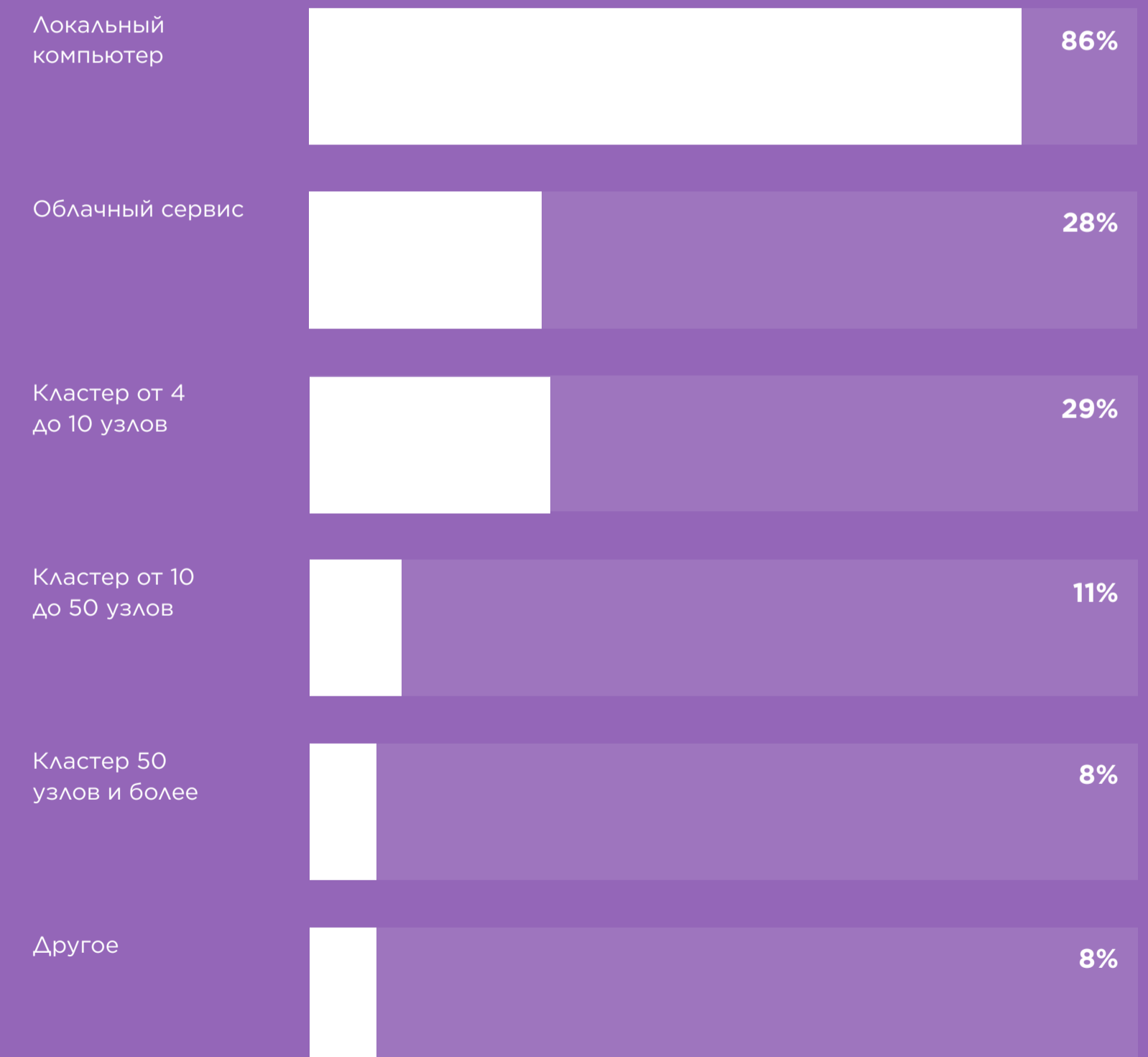
опрошенных используют
Linux для анализа данных

Где Вы производите вычисления?

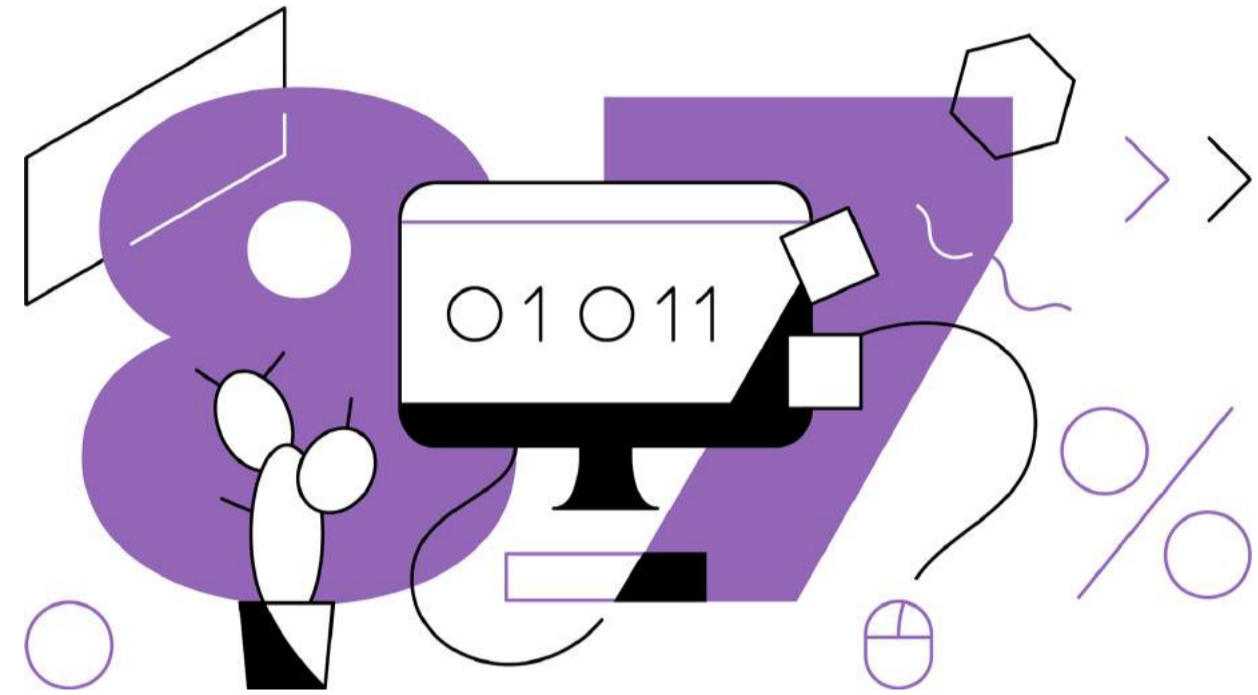
Все

Студенты

Не студенты



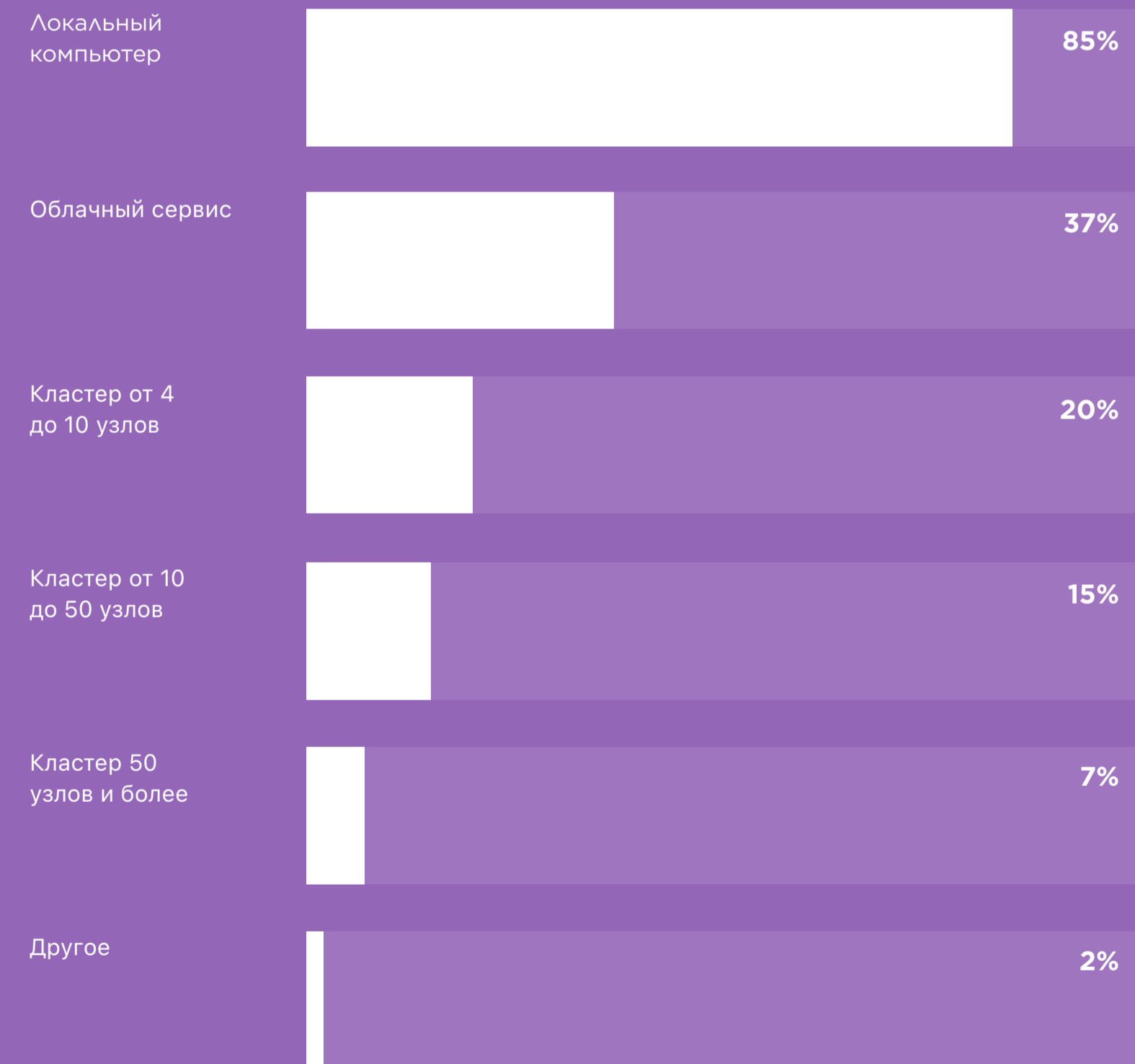
3. Технологии и инструменты



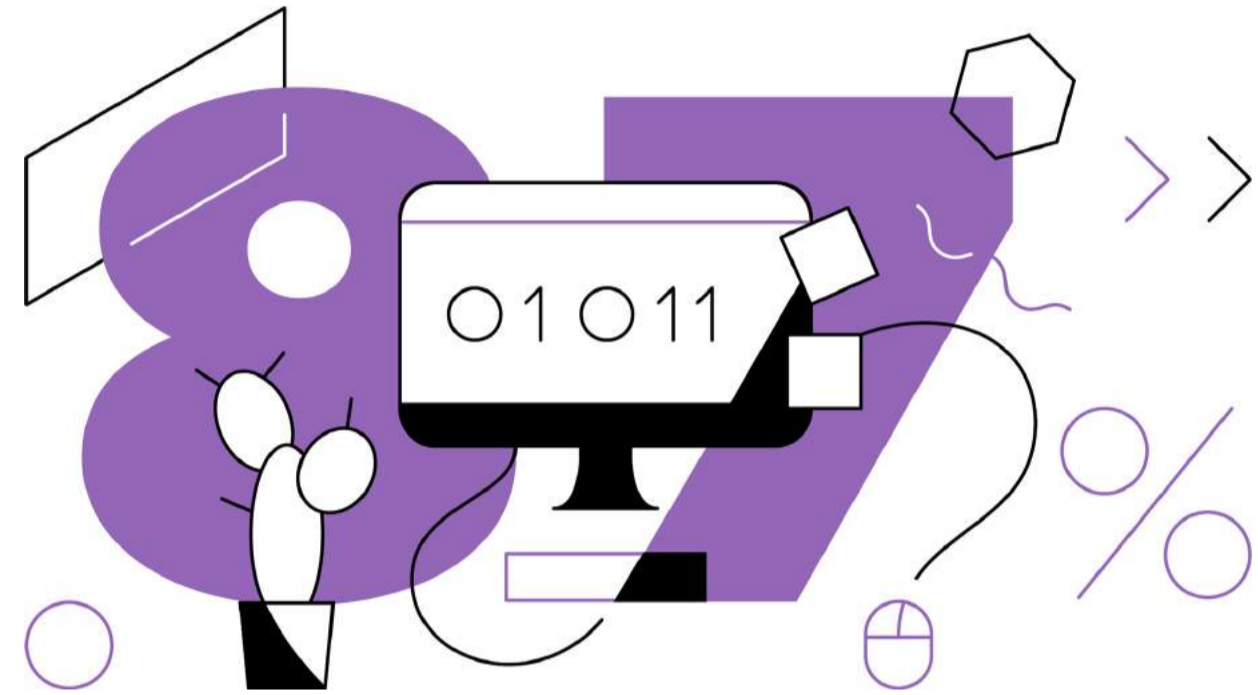
респондентов производят вычисления на локальном компьютере

Где Вы производите вычисления?

Все Студенты Не студенты



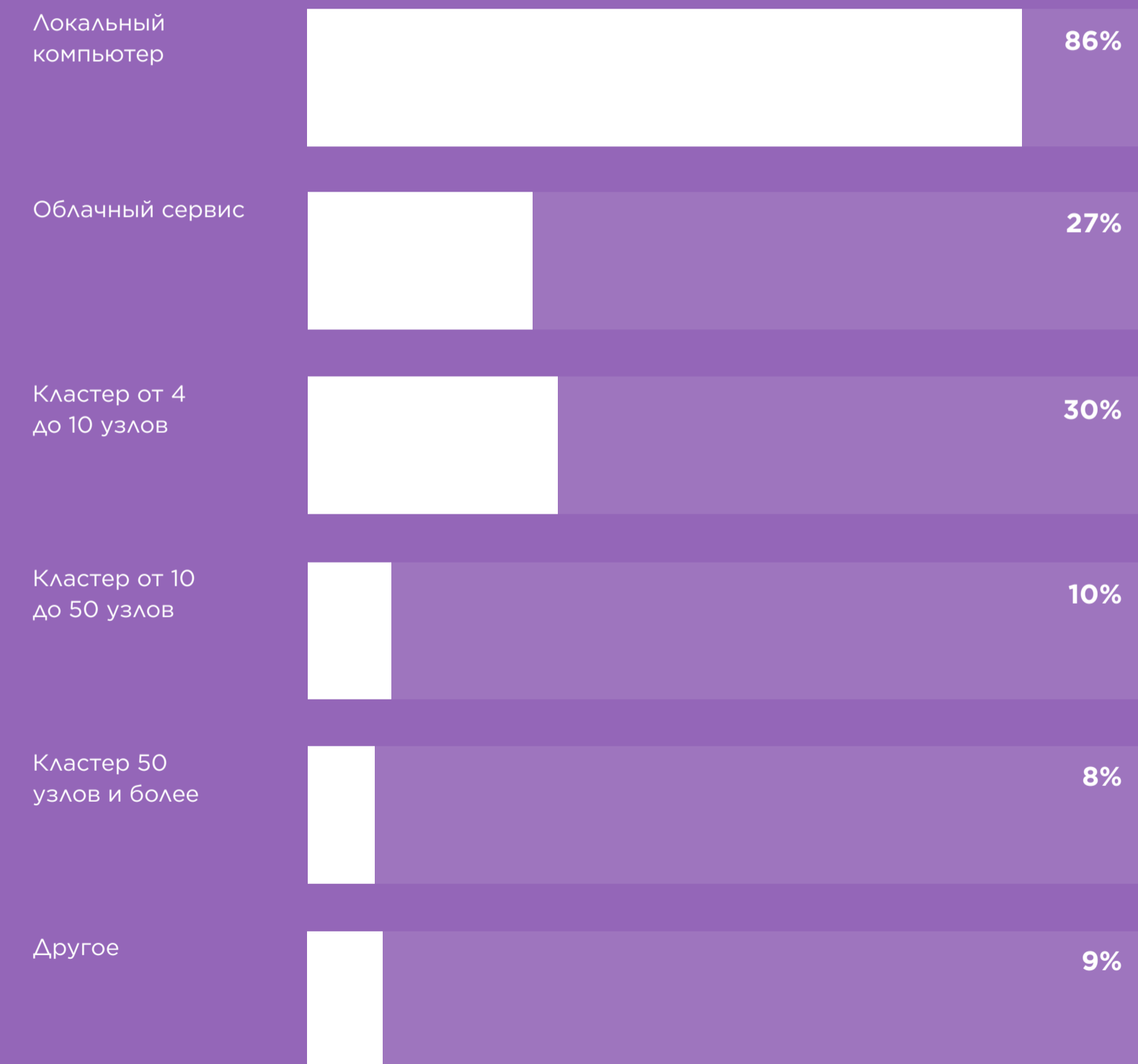
3. Технологии и инструменты



респондентов производят вычисления на локальном компьютере

Где Вы производите вычисления?

Все Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

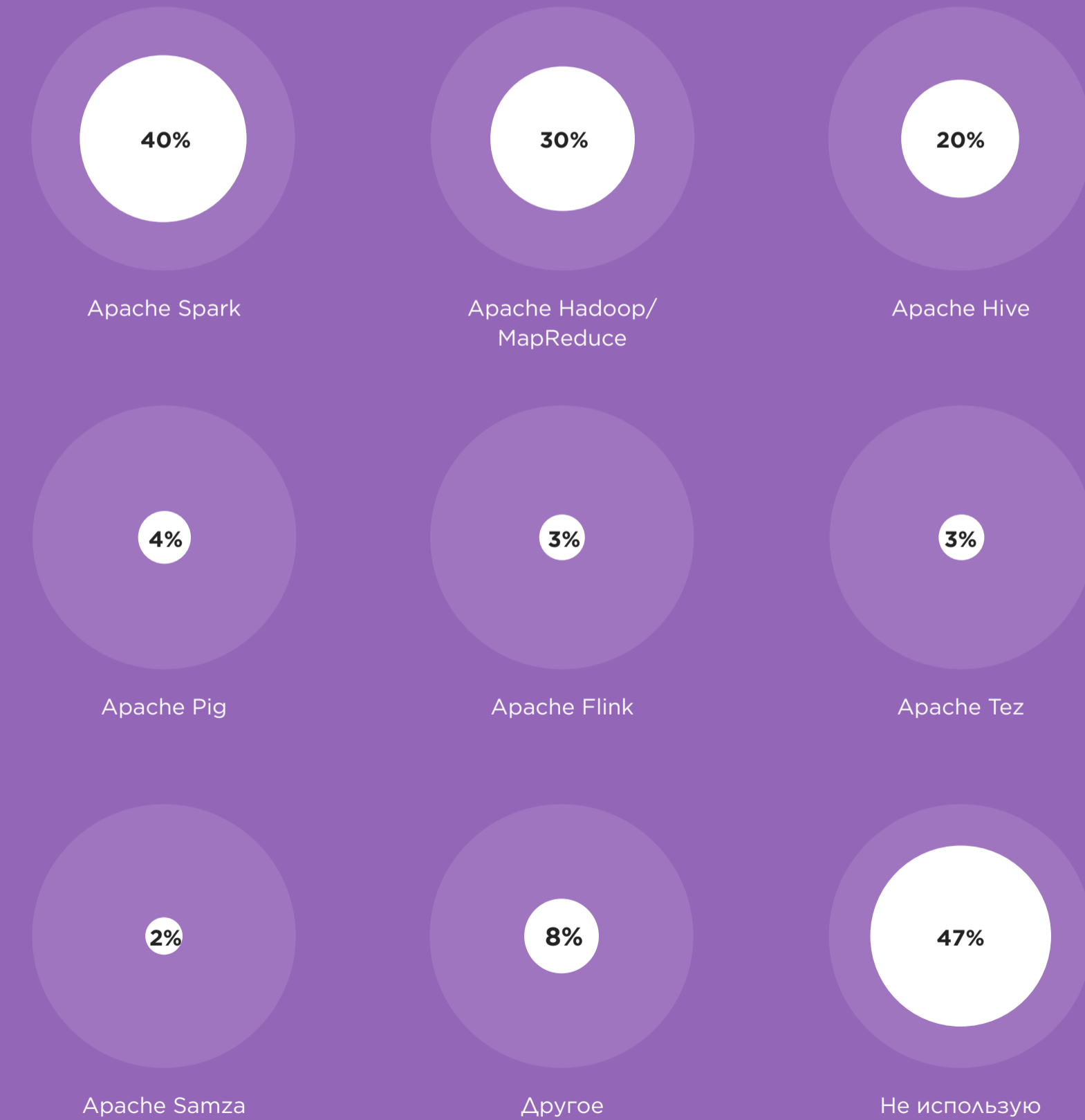
Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data Вы используете?

Все Python R Java Scala Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

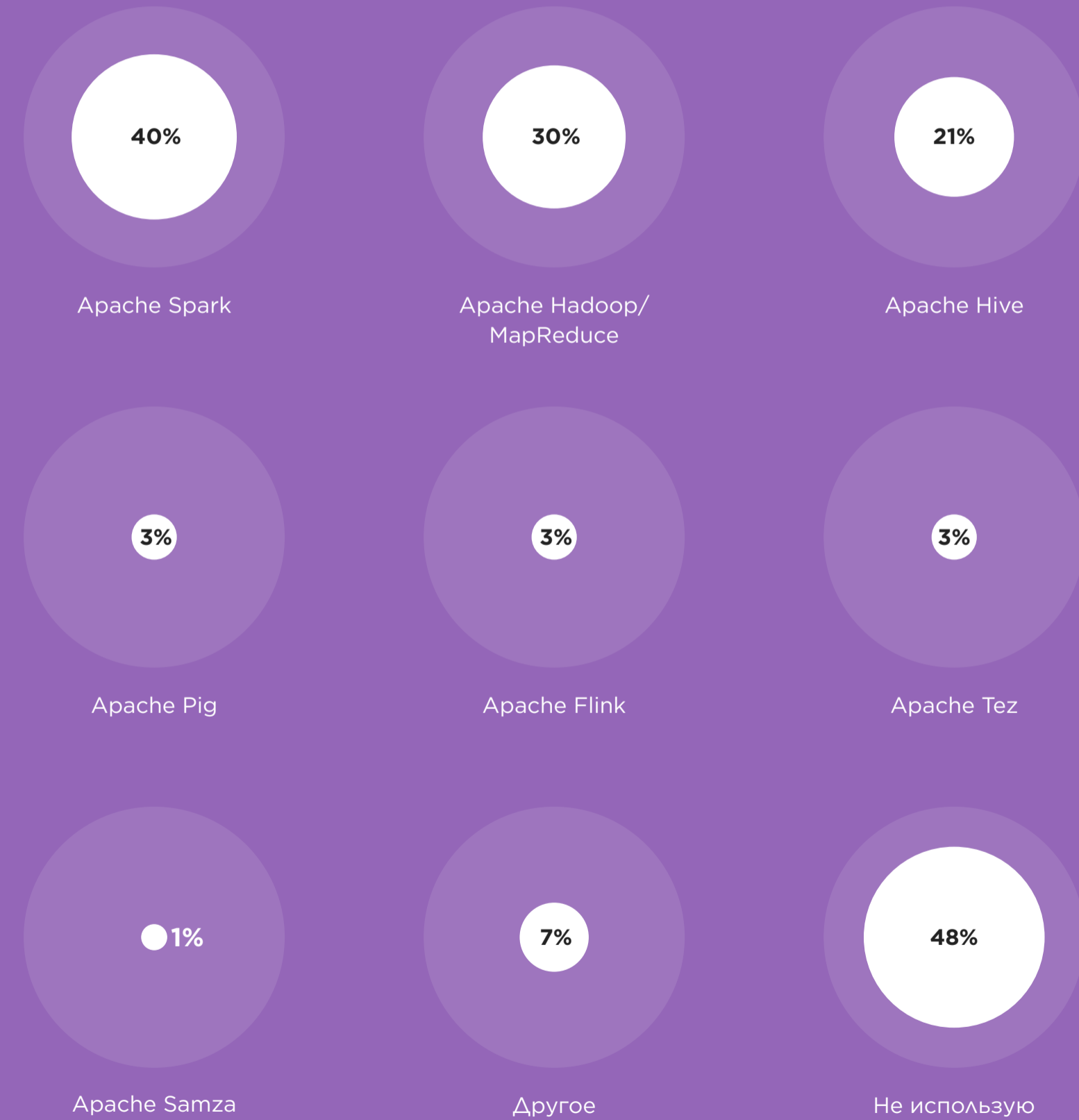
Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data Вы используете?

Все Python R Java Scala Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data Вы используете?

Все Python R Java Scala Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

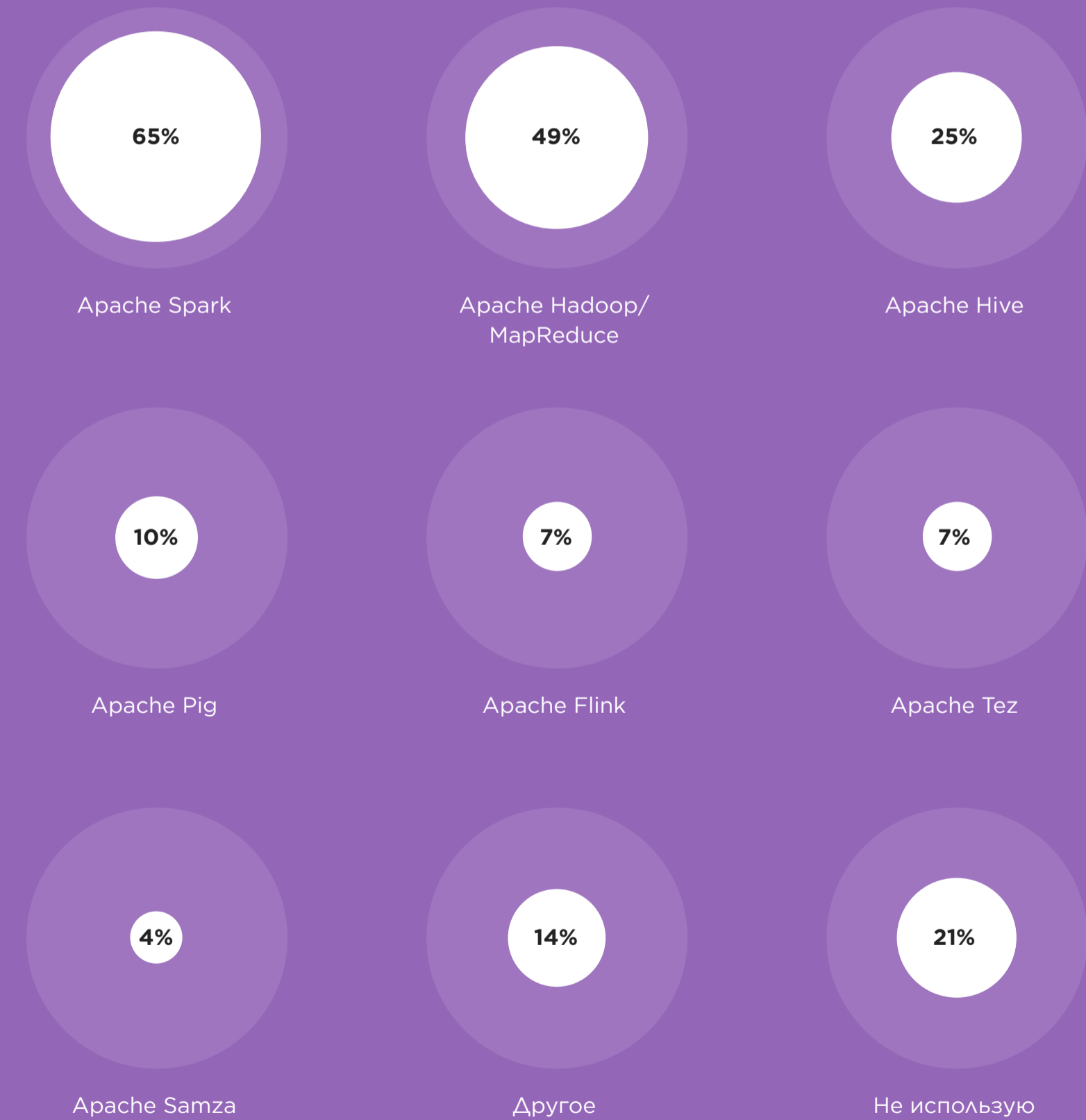
Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data Вы используете?

Все Python R Java Scala Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

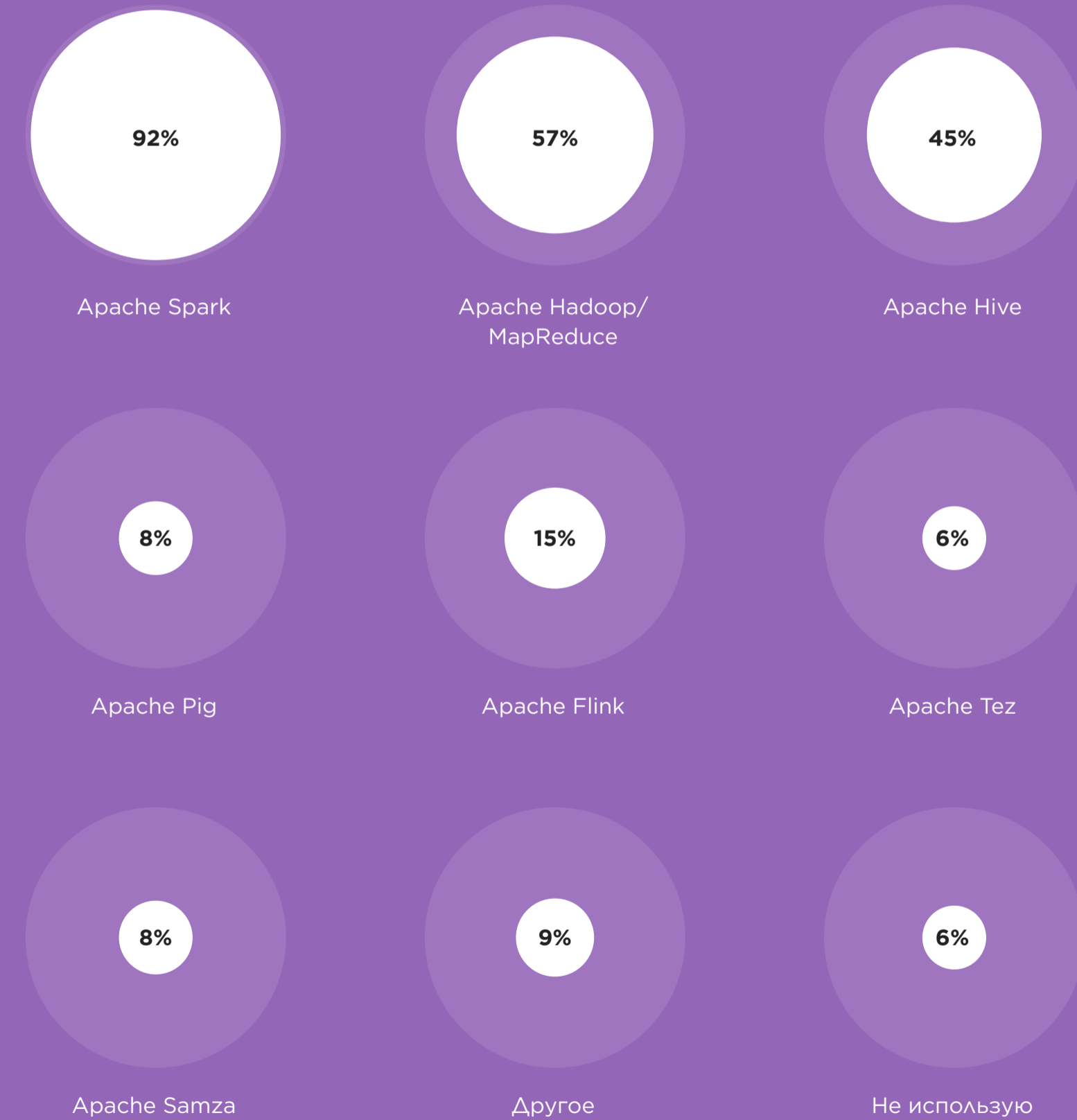
Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data Вы используете?

Все Python R Java Scala Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

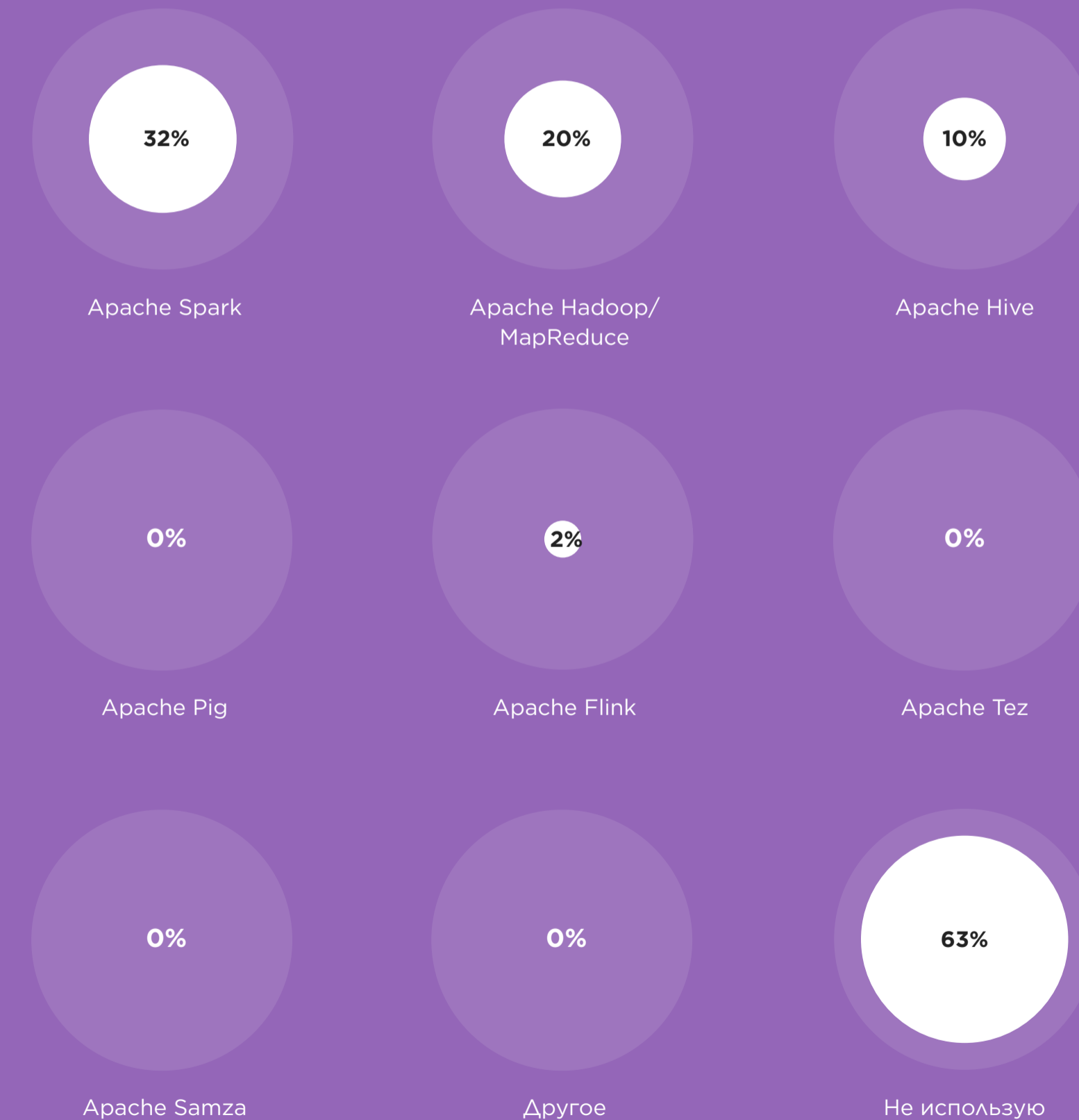
Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data Вы используете?

Все Python R Java Scala Студенты Не студенты



3. Технологии и инструменты

Apache Spark — главный инструмент анализа «больших данных»

Apache Spark используют 40% респондентов, в том числе 92% тех, кто программирует на Scala. Все, для кого Scala является основным языком, используют Apache Spark.

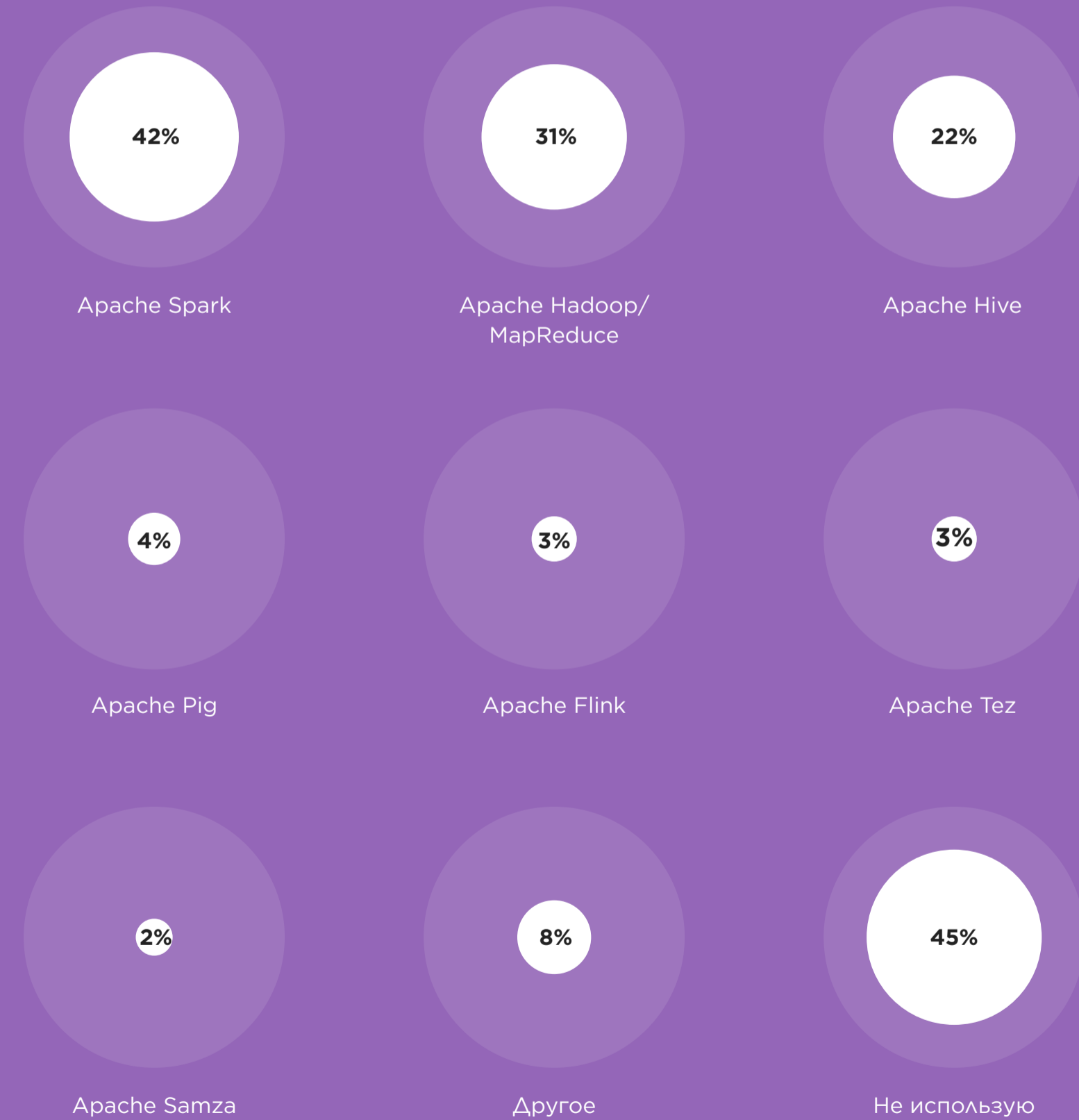
Доля тех, кто программирует только на Python и использует Spark, составляет порядка 14% (если не учитывать возможность использовать Spark из Lua и Julia, то эта доля увеличится до 20%).

Hadoop и Hive несколько менее популярны среди респондентов — 30% и 20%, соответственно.

Почти все респонденты хотя бы иногда пользуются Python, поэтому распределение ответов Python-программистов почти не отличается от общего распределения. Scala же используется в основном только теми, кто занимается Big Data.

Какие инструменты для работы с Big Data вы используете?

Все Python R Java Scala Студенты Не студенты



4. Деньги

Мы не делали никаких предположений о том, что именно подразумевают респонденты под термином “большие данные”, а всего лишь выделили набор технологий, которые наиболее распространены в этой области.

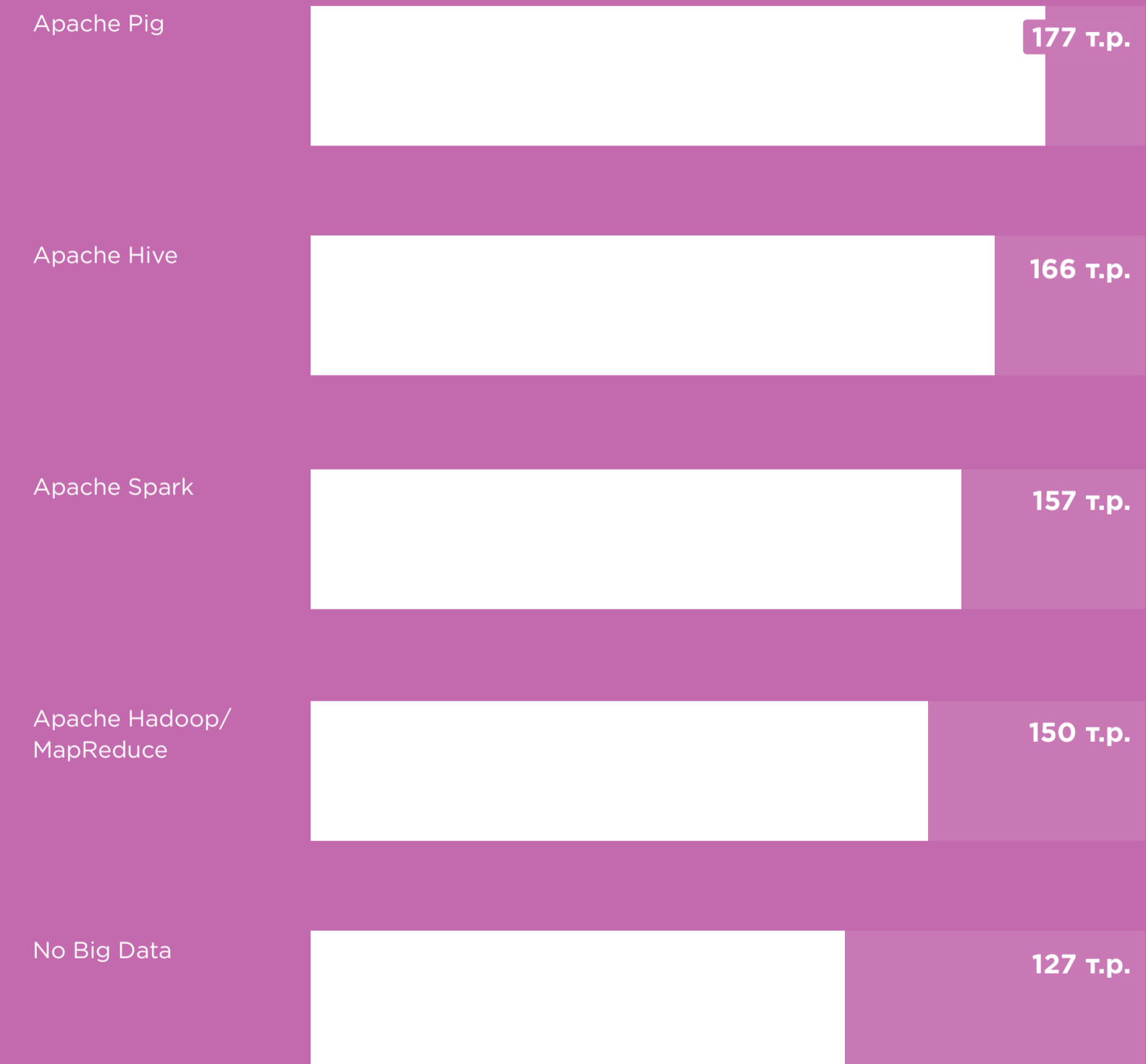
Как и следовало ожидать, знание технологий больших данных — ключ к высокой заработной плате. Средняя зарплата специалиста вне стека технологий больших данных составляет 127 тыс. рублей.

В зависимости от технологий зарплата специалистов в России сильно варьируется, однако в среднем в области анализа больших данных она значительно выше.

При этом, несмотря на популярность Apache Spark, респонденты со знанием этой технологии уступают по зарплате специалистам, владеющим Apache Pig и Apache Hive — 157 тыс. рублей против 177 и 166 тыс. соответственно. Знание Apache Hadoop/MapReduce дает в среднем 150 тыс. рублей.

Это отличается от ситуации в мире, где знание Apache Spark сулит специалисту самую высокую зарплату. Из рассмотрения были исключены Apache Tez и Apache Samza, которыми пользовались менее 10 респондентов.

Средняя сумма заработной платы* Технологии и инструменты.



*Сумма заработной платы, полученной «на руки» в прошлом месяце (в рублях, исключая премии)

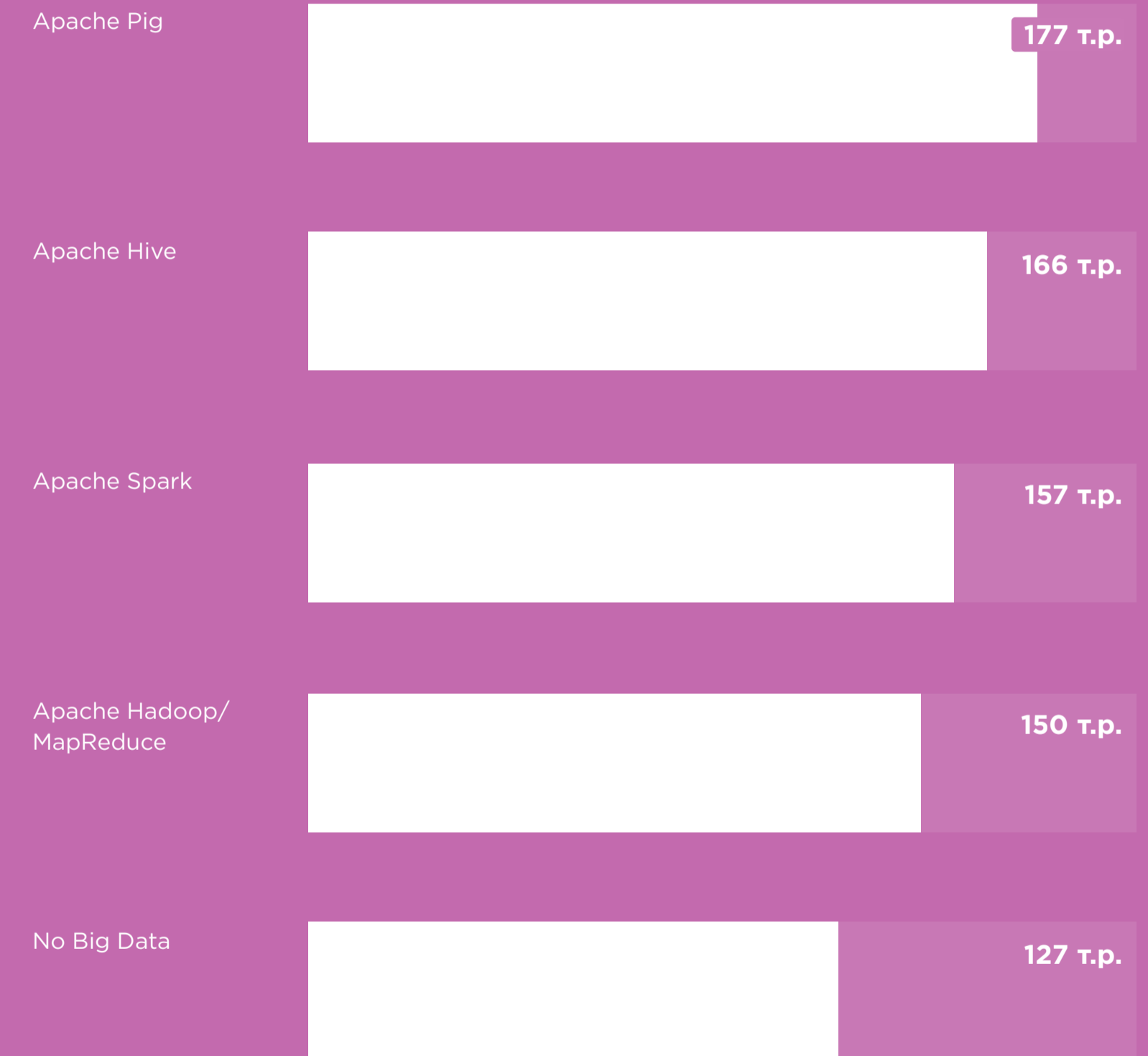
4. Деньги



Дмитрий Бугайченко
инженер-аналитик, OK.RU,
Санкт-Петербург

«Самым интересным показался график про зарплаты. Мне кажется все-таки, что причина такого распределение более глубокая. Зарплата у людей, знающих Pig, выше не потому что они знают Pig, а потому что большинство таких людей относятся к группе специалистов с опытом, плюс должны быть достаточно активны и любознательны, чтобы до него дойти».

Средняя сумма заработной платы* Технологии и инструменты.



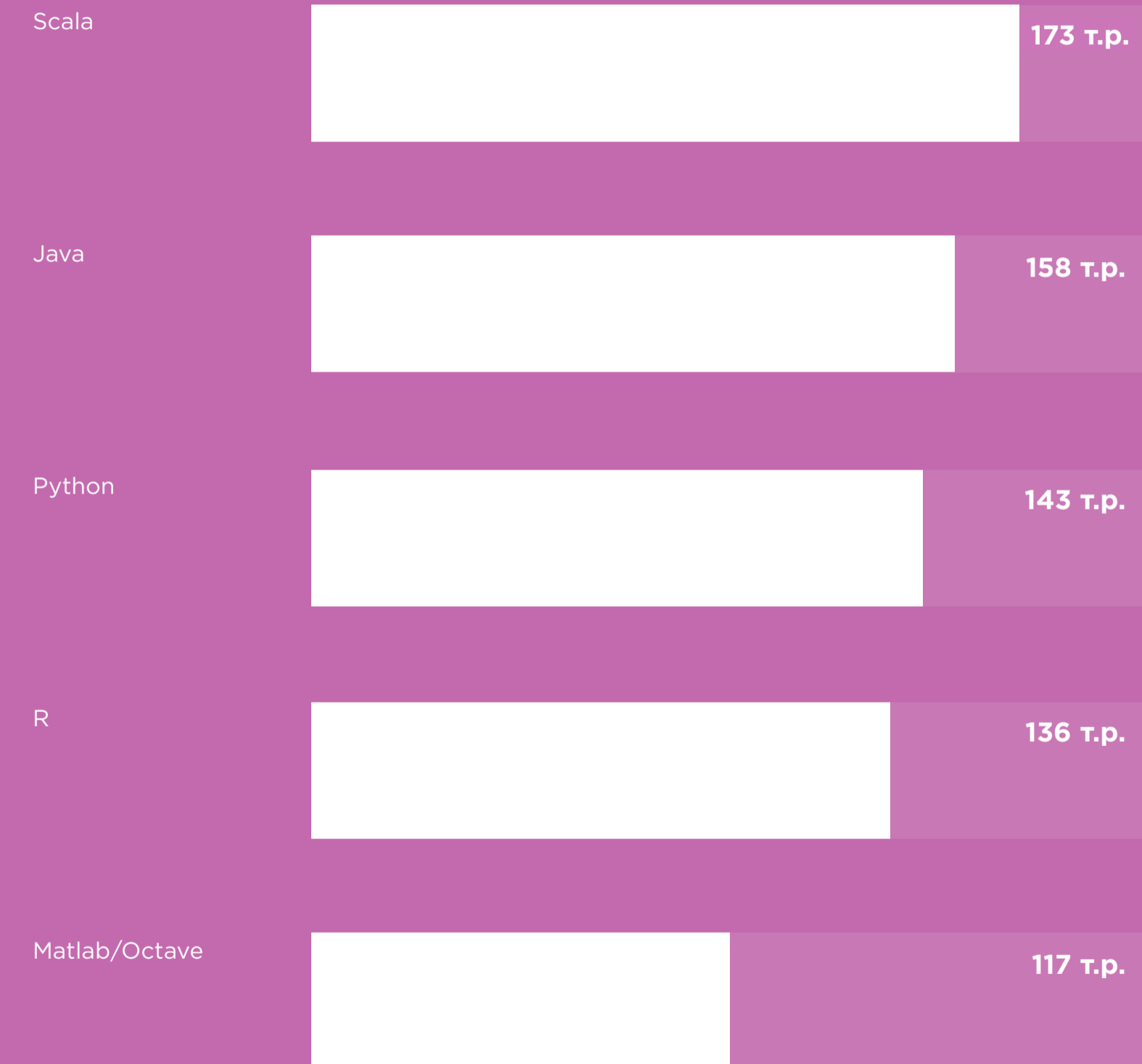
*Сумма заработной платы, полученной «на руки» в прошлом месяце (в рублях, исключая премии)

4. Деньги

В вопросе зависимости зарплат от языка программирования мы не отличаемся от всего мира: специалисты в Scala зарабатывают больше остальных — 173 тыс. в среднем. За ними идут респонденты со знанием Java — 158 тыс., и Python — 143 тыс.

При этом зарплата специалистов, использующих Python, примерно на 4-5% выше, чем специалистов, использующих R (136 тыс.), что вполне соответствует ситуации в мире.

Средняя сумма заработной платы* Языки программирования.



*Сумма заработной платы, полученной «на руки» в прошлом месяце (в рублях, исключая премии)

4. Деньги

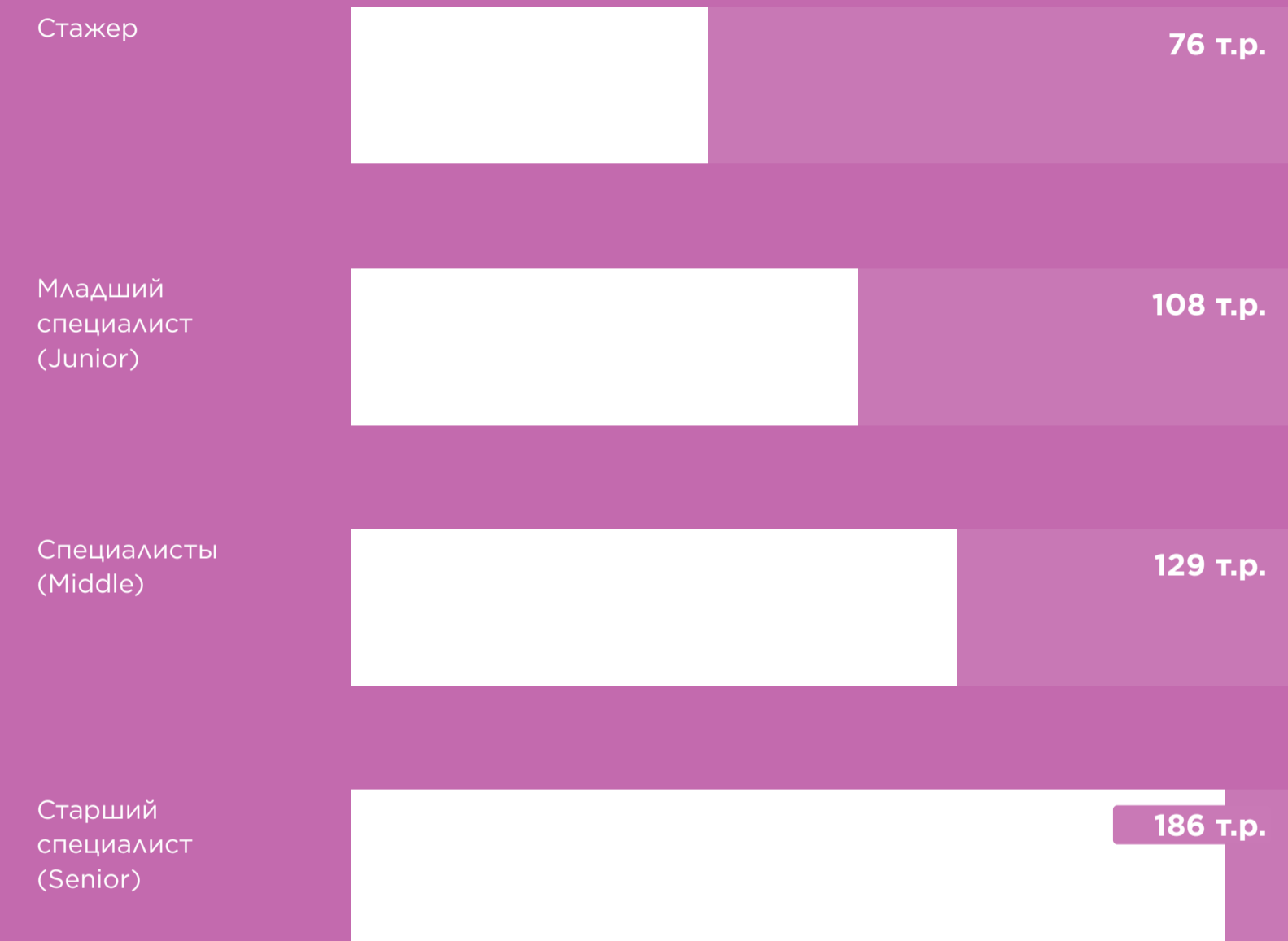
Зарплата также существенно варьируется в зависимости от квалификации и опыта работы.

Что интересно, средняя зарплата стажера составляет 76 тыс. рублей. Младшие специалисты (Junior) получают в среднем 108 тыс., в то время как специалисты (Middle) незначительно больше — 129 тыс. Однако разрыв между старшими специалистами (Senior) и остальными очень существенный, на этом уровне средняя зарплата приближается к 186 тыс.

186
тыс. руб.

в среднем зарабатывает
старший специалист

Средняя сумма заработной платы* Квалификация.



*Сумма заработной платы, полученной «на руки» в прошлом месяце (в рублях, исключая премии)

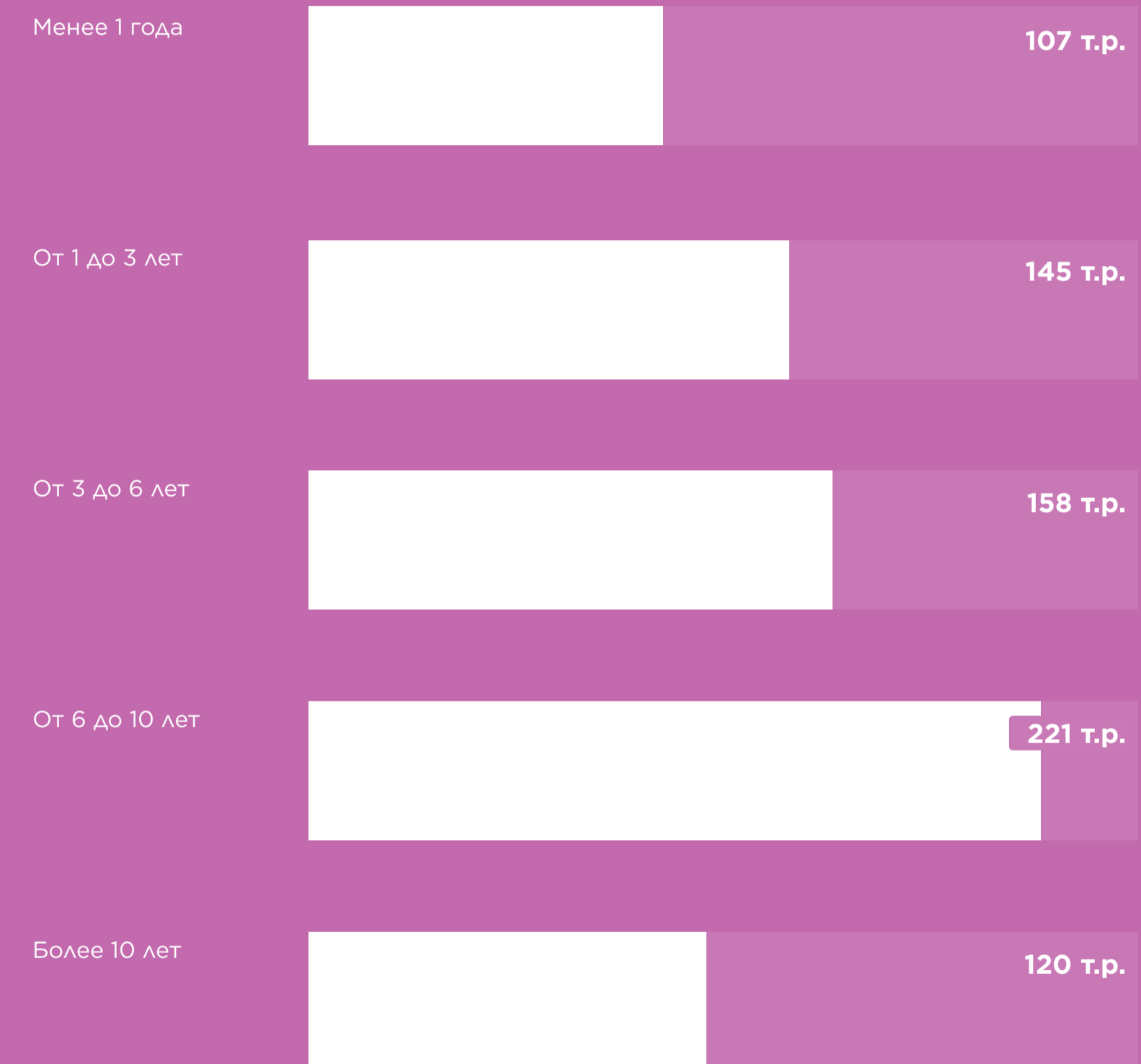
4. Деньги

Ситуация с опытом не такая очевидная, как в случае квалификации. Работники с опытом более 10 лет имеют среднюю зарплату в 120 тыс., что значительно уступает другим категориям, кроме совсем начинающих (с опытом менее 1 года). Дело в том, что респондентов с таким опытом набралось очень мало — всего 8 человек, и 6 из них используют R.

Этому есть вполне понятная причина, ведь язык Python стал популярен среди специалистов по анализу данных менее 10 лет назад, тогда как R занимал абсолютно доминирующую позицию в этой области долгие годы, уступив Python сравнительно недавно.

В остальном с большим отрывом лидируют специалисты со стажем от 6 до 10 лет, зарабатывая 221 тыс. рублей. Респонденты со стажем от 3 до 6 лет получают в среднем 158 тыс. Начинающие могут рассчитывать на 107 тыс., а специалист с опытом от 1 до 3 лет может рассчитывать уже на 145 тыс.

Средняя сумма заработной платы* Опыт работы.



*Сумма заработной платы, полученной «на руки» в прошлом месяце (в рублях, исключая премии)

Спасибо, что уделили нам время!
Мы надеемся, что эта инфографика
оказалась для вас полезна.

Если у вас есть вопросы по данному
исследованию или предложения на будущее,
пишите нам на почту survey@jetbrains.com.

Мы будем рады получить ваши отзывы
и комментарии, чтобы в следующий раз
сделать свою работу еще лучше.