

УДК 311.2

Т.Г. Аналькова, Т.С. Чуприн

Финансовый университет при Правительстве Российской Федерации, Москва,
email: TGApalkova@fa.ru, timopheychuprin@yandex.ru

ВОЗМОЖНОСТИ ЯЗЫКА R В ПОСТРОЕНИИ АНАЛИТИЧЕСКИХ ДАШБОРДОВ НА ПРИМЕРЕ СЕГМЕНТАЦИИ КЛИЕНТСКОЙ БАЗЫ

Ключевые слова: дашборд, визуализация информации о клиентах, язык R, flexdashboard.

Визуализация данных и аналитических отчётов в форме дашбордов получила сегодня широкое распространение в бизнес среде. Помимо специализированных конструкторов, бесплатные версии которых имеют ограниченный функционал, для их построения можно использовать языки программирования с открытым программным кодом, в частности – язык R, в рамках среды R Markdown. Статья преследует цель ознакомить читателя с некоторыми возможностями R в создании дашбордов, поскольку в русскоязычных публикациях этот вопрос практически не освещается. Рассмотренный пример демонстрирует использование различных инструментов библиотек flexdashboard, Shiny и plotly при построении аналитической панели по данным условного примера о клиентах банка. В качестве основных достоинств представленной реализации заявляются: информативность, наличие нескольких тематических панелей-разделов, способность обновляться в режиме реального времени, экономия затрат, объясняемая использованием программного обеспечения с открытым кодом.

T.G. Apal'kova, T.S. Chuprin

Financial University under the Government of the Russian Federation, Moscow,
email: TGApalkova@fa.ru, timopheychuprin@yandex.ru

FEATURES OF R LANGUAGE IN BUILDING ANALYTICAL DASHBOARDS ON THE EXAMPLE OF CUSTOMER BASE SEGMENTATION

Keywords: dashboard, customer information visualization, R, flexdashboard package.

Visualization of data and analytical reports in the form of dashboards has become widespread in the business environment today. In addition to specialized constructors, the free versions of which have limited functionality, you can use open source programming languages, in particular, the R language, within the R Markdown environment to build them. The article aims to acquaint the reader with some of the possibilities of R in creating dashboards, since this issue is practically not covered in Russian-language publications. The considered example demonstrates the use of various tools of the flexdashboard, Shiny and plotly libraries when building an analytical panel based on the data of a conditional example about bank customers. The main advantages of the presented implementation are: information content, the presence of several thematic panels-sections, the ability to update in real time, cost savings due to the use of open source software.

Тема создания аналитических панелей, так называемых дашбордов, инструментов объединяющих современные технологии с многолетним опытом графического представления информации, уже не нова, но продолжает оставаться актуальной. Дашборд – это удобный и эффективный инструмент представления информации о бизнесе, который может быть настроен с учётом специфики адресной аудитории – руководителей, адресантов маркетинговой компании, клиентов, контрагентов, потенциальных инвесторов. Курсы подготовки IT специалистов предлагают программы, ориентированные на работу со специализированными конструкторами дашбордов, среди наиболее популярных Tableau, Ms Power BI, Google Looker. Однако важнейшие свой-

ства качественной аналитической панели – ее гибкость, возможность настройки с учётом специфик отражаемого процесса и целевой аудитории. По этим характеристикам с программами – конструкторами успешно конкурируют современные языки программирования. Обзор источников показывает, что русскоязычные публикации практически не раскрывают тему применения языка R для создания дашбордов, хотя этот язык с открытым кодом представляет широчайшую базу инструментов, предназначенных для решения этой задачи. В статье предпринимается попытка в некоторой степени заполнить этот информационный пробел, представив несколько библиотек языка R, реализующих возможность построения аналитических панелей.

Цель исследования

Уровень развития материально-технической базы и инструментальных средств анализа и визуализации данных позволяет сегодня создавать сколь угодно сложные и информативные аналитические панели. Многочисленные обзоры представляют он-лайн и десктопные версии программных продуктов для создания дашбордов, выбор пользователя при этом зависит от предметной области, бюджета, сложности анализа, технических ограничений (например, тип операционной системы). Однако на наш взгляд в русскоязычном сегменте статей и блогов, представляющих эти программные средства, практически не раскрываются возможности создания аналитических панелей средствами языка программирования R. Настоящая статья преследует цель заполнить этот пробел, продемонстрировав на примере применения среды разработки R Markdown, библиотек flexdashboard и Shiny эффективность R в создании интерактивных инструментов визуализации.

Материал и методы исследования

Графическое представление данных, или визуализацию данных вполне можно отнести к одному из неотъемлемых элементов предварительной обработки данных. На этапе формирования отчёта правильно построенные и скомпонованные диаграммы позволяют сделать информацию более выразительной, удобной для восприятия и запоминающейся. Отмечено, что «графический метод <...> является продолжением табличного метода. То, что при чтении таблицы может остаться незамеченным, обнаруживается на графике» [1, стр 62]. Опыт докладов о результатах аналитических исследований дает основания согласиться с тезисом В.В. Платонова о том, что при коммуникациях «человек (аналитик) – человек (принимающий управленческое решение)» информация, представляя собой итоговый результат постижения сложного хозяйственного объекта или экономического явления, должна быть выражена наиболее наглядным визуальным образом [2]. Нельзя также забывать, что представление информации в графической форме позволяет донести идеи докладчика в максимально короткие

сроки, что зачастую крайне актуально при коммуникации с руководителем.

Что касается методов визуализации, наверняка, существует несколько подходов к их классификации, один из них раскрыт в работе [3]. В настоящей статье рассматривается метод составной визуализации (compound visualization), как наиболее актуальный при презентации сложных экономических явлений и объектов. Обычно этот метод подразумевает взаимодополняющее использование различных форматов графического представления в одной схеме или фрейме [3]. Типы используемых диаграмм при этом должны соответствовать типу сравнения данных. Джин Желязны в Пособии по визуальным коммуникациям выделяет 5 типов сравнений и соответствующих им диаграмм [4]: покомпонентное, позиционное, временное, частотное и корреляционное сравнение. Возможно, этот перечень не исчерпывающий, например, не лишним было бы добавить пространственное сравнение. Такая типология диаграмм отвечает большинству решаемых при создании экономических отчётов задач. Для визуализации в рамках первичного анализа может быть использована другая классификация, поскольку он решает иные задачи.

Остановившись на методе составной визуализации, рассмотрим один из наиболее удобных современных ее инструментов – создание дашбордов (от англ. dashboard). Под дашбордом (иначе – дэшбордом) в контексте визуализации данных понимается интерактивная аналитическая панель, иными словами – электронный документ, содержащий в сжатой форме наиболее важные сведения отчёта и графики, реализуемый, как правило, в форме веб-страницы.

На сегодняшний день не существует единой устоявшейся классификации дашбордов, однако по уровню целеполагания можно выделить операционные, стратегические и тактические дашборды, по предметной области – финансовые, маркетинговые и т.п. Иногда в качестве отдельного вида дашборда выделяют так называемые «информационные массы», которые не содержат выводов и рекомендаций, но отражают некоторые тенденции. Адресанты такого дашборда в состоянии самостоятельно извлечь

информацию из отображенных на дашборде сведений. Инфомассам противопоставляют аналитические дашборды, содержащие результаты анализа и основанные на них рекомендации. Специалисты Яндекс-Практикума отмечают, что основное отличие дашборда от инфографики заключается в том, что последняя подходит только для упрощенного отражения информации, основное отличие от отчёта – интерактивность дашборда и предопределенная возможность отображать изменения в динамике [5].

Существует множество инструментальных средств создания дашбордов, отличающихся стоимостью, возможностями масштабирования и доступной глубиной аналитики. Чтобы опубликовать группу связанных графиков в виде информационной панели, можно использовать программную среду R Markdown, которая сочетает в себе возможности редакторов текста и программного кода и изначально создавалась для удобного оформления результатов в области data science. В общем случае среда R Markdown позволяет использовать фрагменты программного кода, написанного на R, Python и SQL и поддерживает множество статических и динамических форматов вывода, в том числе PDF, Ms Word, html.

Рассмотрим далее возможности библиотеки flexdashboard в построении дашбордов на языке с открытым программным кодом R. R обладает несомненными преимуществами благодаря широкому спектру базовых статистических

функций, а также специальных библиотек, реализующих возможности систематизации, аналитики данных и их графического представления. Свойство гибкости, отраженное в названии библиотеки (flexdashboard от flexible dashboard) заключается в разнообразии возможностей расположения компонентов и организации страниц дашборда.

Так, например, при желании расположить диаграммы в двух колонках, как показано на рисунке 1, применяется следующий код, созданный в формате R Markdown:

```

---
title: "Макет 1"
output: flexdashboard::flex_dashboard
---
Column
-----
### Диаграмма 1
```{r}
```

Column
-----
### Диаграмма 2
```{r}
```

### Диаграмма 3
```{r}
```

```

Начало столбца обозначается записью Column и множественным дефисом под ней.

В результате макет панели имеет вид:

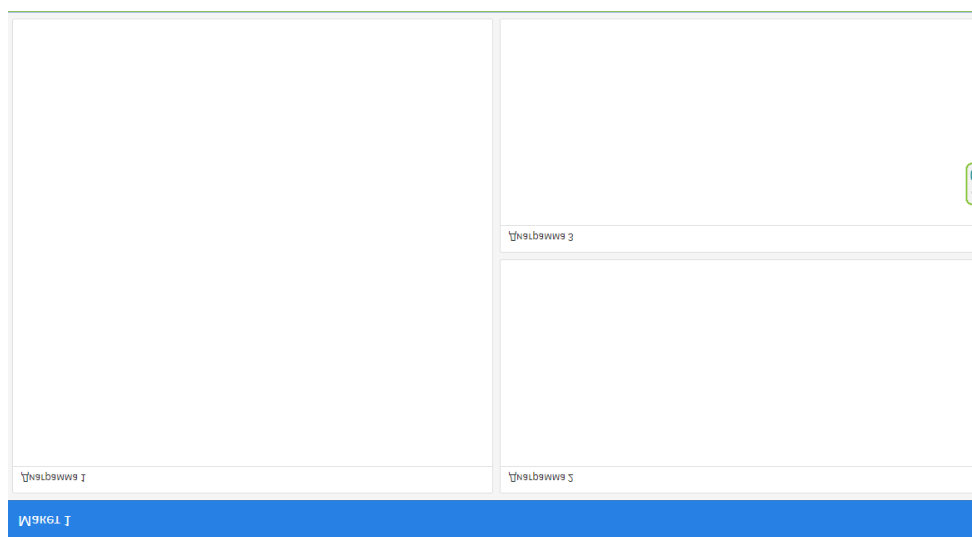


Рис. 1. Макет дашборда при позиционировании графиков по столбцам

Несколько компонентов могут быть видны одновременно, располагаясь друг под другом, или рядом. Если нет необходимости делать одновременно видимыми все объекты, некоторые из них можно разместить на временно неактивных вкладках, добавляя к заголовку столбца `{.tabset}`, или `{.tabset .tabset-fade}`.

Для использования совместно с `flexdashboard` доступны несколько библиотек, реализующих JavaScript визуализацию данных (проще говоря – динамическую визуализацию): `Leaflet`, `dygraphs`, `Plotly`, `rbokeh`, `Highcharter`, `visNetwork`. Создаваемые диаграммы при этом динамически меняют размер и всегда вписываются в рамки предназначенных для них областей. Создание динамических дашбордов доступно с использованием библиотеки `Shiny`, однако не является обязательным. В следующем ниже примере будут рассмотрены эти и некоторые другие возможности создания дашборда на языке R с использованием `flexdashboard`.

Результаты исследования и их обсуждение

Процесс спецификации дашборда рассмотрим далее на условном примере, набор данных для которого взят из базы `Kaggle` [6]. Упомянутый набор данных содержит сведения о 5000 клиентах – держателях вкладов некоторого коммерческого банка. Каждый клиент характеризуется рядом признаков – доход, размер семьи, опыт профессиональной деятельности и прочими. Особый интерес для банка представляет возможность конверсии так называемых «пассивных» клиентов в категорию заёмщиков, при условии сохранения их в качестве вкладчиков. Среди пяти тысяч клиентов, чьи сведения отображены в наборе данных, 480 перешли в категорию заёмщиков во время последней маркетинговой кампании. Детальный анализ особенностей этих клиентов может позволить сделать подобные маркетинговые акции более адресными, ориентированными на конкретные группы вкладчиков. Таким образом, цель дашборда – продемонстрировать специфические черты клиента – потенциального заёмщика.

Предлагаемый в качестве примера реализации технологии дашборд

состоит из нескольких панелей вкладок, одна из которых всегда активна. Каждая панель содержит информацию о клиентах банка: личного характера, сведения, характеризующие их как пользователей банковских услуг, специфическую информацию, связанную с получением потребительского кредита во время последней маркетинговой акции. На рисунке 2 показан вид панели «Личная информация». Панель содержит два основных элемента: динамические диаграммы, которые отражают сегментацию всех клиентов по возрасту, семейному положению, стажу работы, уровню образования и доходам и выделенные контрастным цветом блоки, так называемые `Value Boxes`, каждый из которых содержит только одно значение – среднее по показателю, или общее количество. Так видно, что по уровню образования и семейному положению все клиенты распределяются равномерно, практически равномерно – по стажу (за исключением единичных случаев нулевого стажа и стажа, превышающего 40 лет). Значительное большинство клиентов (более 2000) имеют доход в диапазоне от 30 до 80 тысяч долларов в год, более 3000 клиентов находятся в возрасте от 31 до 59 лет.

Каждая диаграмма представляет собой динамический график, при наведении курсора на элемент, на экране появляется его значение, также можно добавлять, или исключать из графика категории. Эти свойства достигаются при помощи использования для построения графиков функций библиотеки динамической визуализации `plotly`: `plot_ly()` задает динамическую диаграмму как тип, внешние по отношению к ней функции `add_pie()` и `add_bars()` определяют тип графика.

Элементы `Value Boxes` позволяют акцентировать внимание целевой аудитории дашборда на отдельных ключевых значениях. Они задаются функцией `flexdashboard::valueBox()`, в качестве аргументов выступают: текст, значение показателя, цвет, пиктограмма из стандартного набора библиотеки. Более интересный пример реализации `Value Boxes` демонстрируется на панели «Банковская информация», рисунок 3.

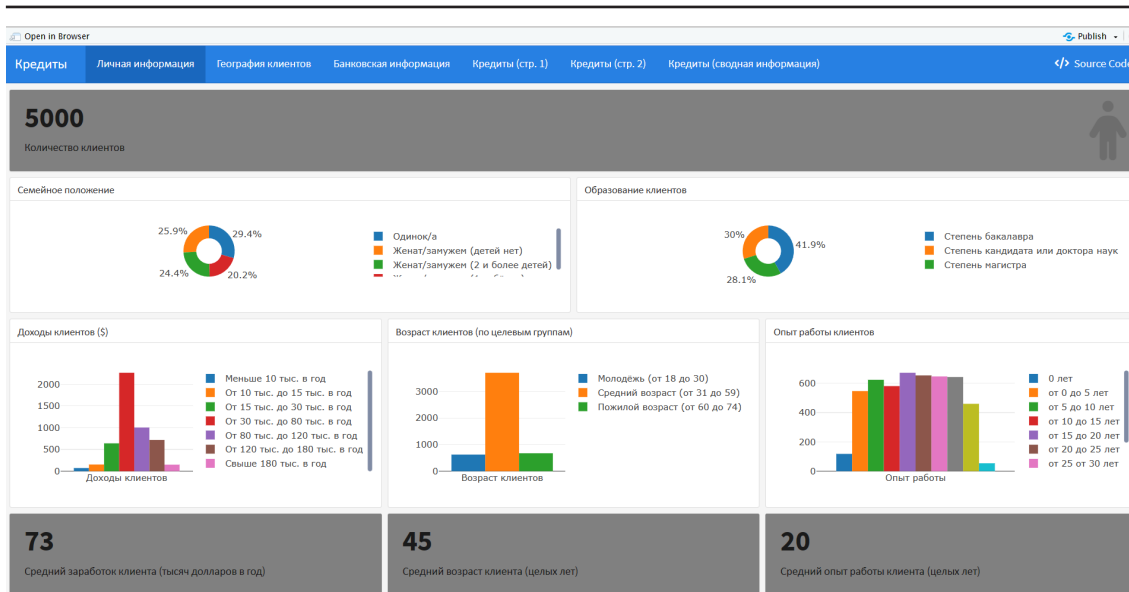


Рис. 2. Общий вид панели «Личная информация»

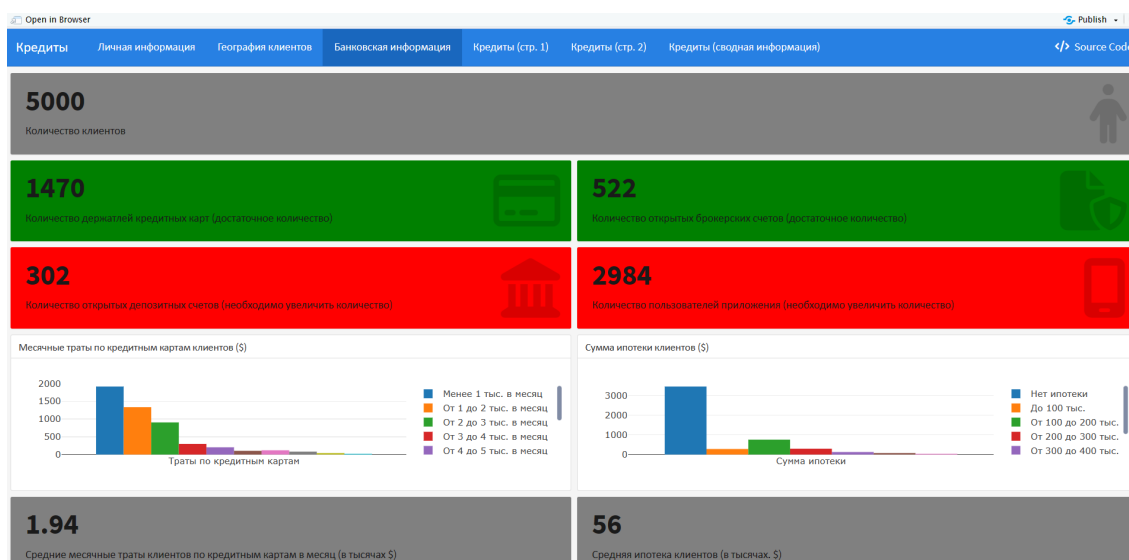


Рис. 3. Панель «Банковская информация»

Предусмотрено, что элементы типа Value Boxes меняют цвет по мере обновления исходных данных, в зависимости от того, удовлетворяет ли значение показателя внутреннему нормативу. Эта возможность доступна благодаря использованию в коде условного оператора if else. Ниже демонстрируется фрагмент кода для Value Box «количество пользователей приложения»:

```
if(quantity_online > quantity_clients*0.75) {
```

```
    valueBox(value = quantity_online,icon = "fa-solid fa-mobile-screen",caption = "Количество пользователей приложения (достаточное количество)",color = "green")
```

```
  } else {
    valueBox(value = quantity_online,icon = "fa-solid fa-mobile-screen",caption = "Количество пользователей приложения (необходимо увеличить количество)",color = "red")
  }
```

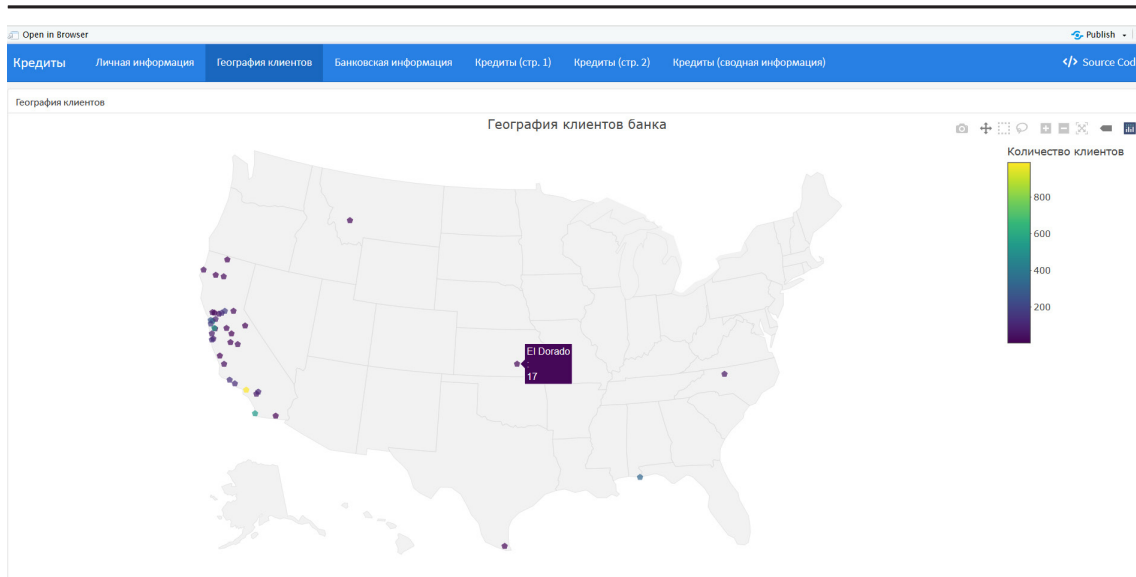


Рис. 4. Карта «География клиентов»

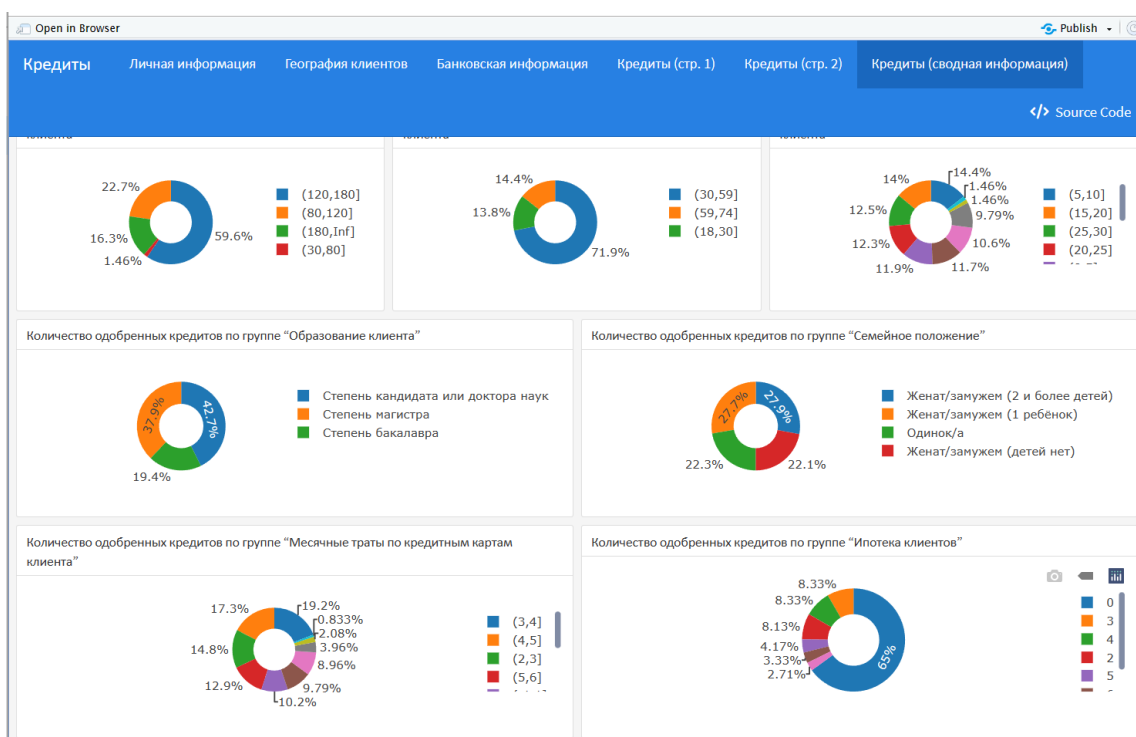


Рис. 5. Информация о клиентах, получивших потребительский кредит

Если доля клиентов, использующих мобильное приложение, оказывается ниже 75%, элемент сигнализирует об этом красным цветом, в противном случае цвет фона зелёный. Таким образом, несмотря на внешнюю простоту, элементы Value Boxes, могут быть

очень информативными и полезными, поскольку акцентируют внимание адресанта дашборда на наиболее важных обстоятельствах и, возможно, проблемах.

Для любой компании важна география клиентов. Дашборд – хороший способ отразить ее в виде карты, рисунок 4.

На картограмме видно, что большинство клиентов проживают в городах штата Калифорния, максимальное количество – в Лос-Анджелесе.

Картограмма также динамическая, она масштабируется и при наведении курсора появляется всплывающая подсказка, показывающая точное число клиентов в конкретном городе. Эти эффекты достигаются использованием функции `plotly::plot_geo()`.

Наконец, на последней панели сделан акцент на информацию о категориях клиентов, получивших персональный кредит во время последней маркетинговой кампании, рисунок 5. Круговые диаграммы на этой панели показывают особенности клиентов, которые согласились на получение потребительского кредита. Так видно, что наибольшая доля одобренных потребительских кредитов приходится на клиентов с высоким (более 120 тысяч долларов в год) доходом возраста от 30 до 59 лет, не имеющих ипотечных кредитов. Кроме того, больший процент получивших потребительский кредит – клиенты с детьми, имеющие уровень образования не ниже степени магистра. Таким образом, клиентов – получателей потребительских кредитов главным образом выделяет высокий уровень дохода и отсутствие ипотеки.

Эта информация позволит определить целевые группы следующей маркетинговой акции и сделать рекламу более адресной.

Выводы

Визуализация данных – инструмент, позволяющий донести информацию в более короткие сроки по сравнению с текстом, или таблицами, сделать ее более запоминающейся и в некоторых

случаях – более понятной. Сложные процессы и явления, которые изучает экономика, многогранны и требуют разностороннего отображения. Кроме того, для принятия адекватных управленческих решений «под рукой» руководителя должны быть всегда актуальные сведения, следовательно, графики должны обновляться в режиме реального времени. Всеми указанными свойствами обладают дашборды, интерактивные аналитические панели.

Приведенный пример демонстрирует, во-первых, удобство и функциональность дашборда как инструмента презентации информации в визуальной форме. Несомненным достоинством при этом является реализованная возможность систематизировать информацию при помощи отдельных тематических панелей. Во-вторых, создание подобного дашборда потребует определенных временных ресурсов, но доступно практически любому аналитику, работающему в среде R без дополнительных финансовых вложений. Фактически, всё, что может потребоваться – это знание функционала языка и наличие компьютера с доступом в Интернет, поскольку даже при отсутствии установленного специального программного обеспечения работа в R возможна с использованием облачных технологий `posit.cloud`.

Наконец, созданный один раз дашборд, с одной стороны, не нуждается в корректировке при обновлении данных на сервере, где хранится информация о клиентах, с другой стороны, при необходимости отобразить какие-либо неучтенные ранее аспекты, исходный код может быть дописан и требуемые недостающие элементы будут отображены.

Библиографический список

1. Федоров М., Спирин А., Башина О. и др. Общая теория статистики / Под общей редакцией А. Спирина, О. Башиной. М.: Финансы и статистика, 1996. 296 с.
2. Платонов В.В. Визуализация больших данных в экономических науках в условиях информационного общества // Вопросы инновационной экономики. 2020. № 4. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/vizualizatsiya-bolshih-dannyh-v-ekonomicheskikh-naukah-v-usloviyah-informatsionnogo-obschestva> (дата обращения: 08.07.2023).
3. Towards A Periodic Table of Visualization Methods for Management // Портал `visual-literacy.org`. [Электронный ресурс]. URL: https://www.visual-literacy.org/periodic_table/periodic_table.pdf (дата обращения: 10.07.2023).

4. Желязны Д. Говори на языке диаграмм: Пособие по визуальным коммуникациям для руководителей / Пер. с англ. М.: Институт комплексных стратегических исследований, 2004. 220 с.
5. Дашборд: что это, типы дашбордов – инструменты и правила создания дашбордов, примеры настройки // Блог Яндекс Практикума. [Электронный ресурс]. URL: <https://practicum.yandex.ru/blog/что-такое-dashbord/> (дата обращения: 10.07.2023).
6. Набор данных Personal Loan Modeling. [Электронный ресурс]. URL: <https://www.kaggle.com/datasets/teertha/personal-loan-modeling> (дата обращения: 15.07.2023).