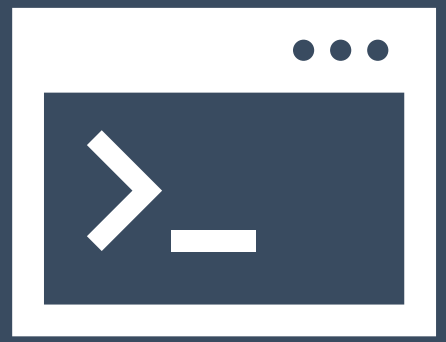
HBC
Harvard Chan Bioinformatics Core

Introduction to R

https://tinyurl.com/hbc-r-revised



Harvard Chan Bioinformatics Core

Introductions!

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth Partan

Emma Berdan

James Billingsley

Zhu Zhuo

Maria Simoneau

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth Partan

Emma Berdan

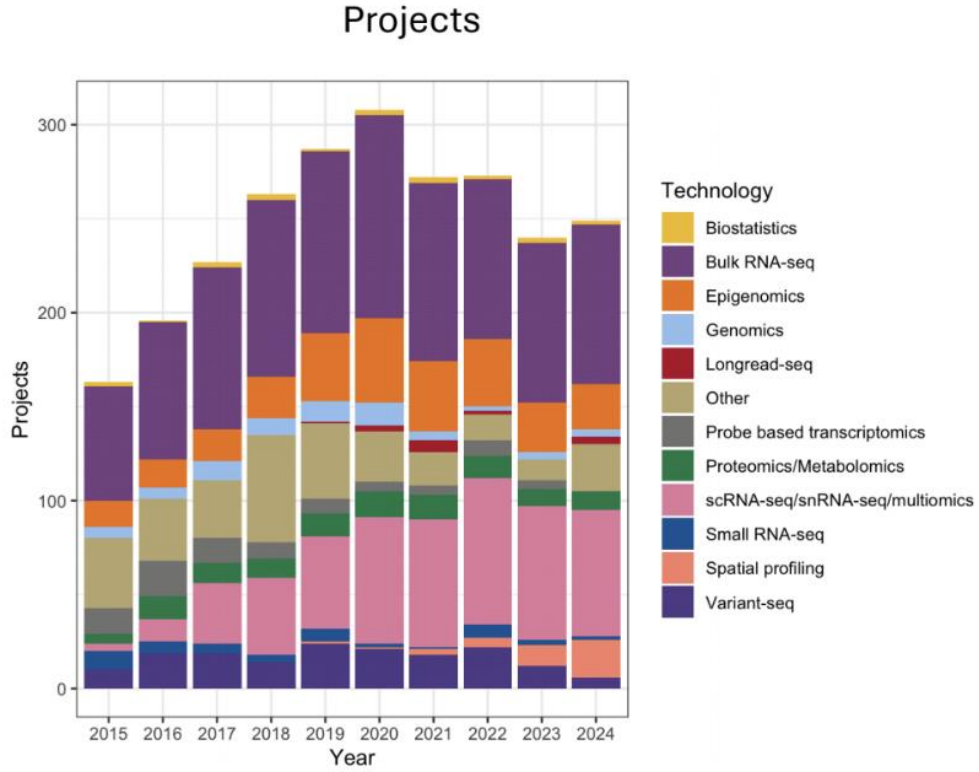
James Billingsley

Zhu Zhuo

Maria Simoneau

Consulting

- Transcriptomics: Bulk, single cell, small RNA
- Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- Variant discovery: WGS, resequencing, exome-seq and CNV
- Multiomics integration
- Spatial biology
- Experimental design and grant support



https://bioinformatics.sph.harvard.edu/services

Consulting

- Transcriptomics: Bulk, single cell, small RNA
- Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- Variant discovery: WGS, resequencing, exome-seq and CNV
- Multiomics integration
- Spatial biology
- Experimental design and grant support

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

Training

- Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
  - Basic Data Skills
    - Shell
    - R
  - Advanced Topics: Analysis of high-throughput sequencing data
    - Chromatin Biology
    - Bulk RNA-seq
    - Differential Gene Expression
    - scRNA-seq
    - Variant Calling
  - Current Topics in Bioinformatics

https://bioinformatics.sph.harvard.edu/training

Training

- Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
  - Basic Data Skills
    - Shell
    - R
  - Advanced Topics: Analysis of high-throughput sequencing data
    - Chromatin Biology
    - Bulk RNA-seq
    - Differential Gene Expression
    - scRNA-seq
    - Variant Calling
  - Current Topics in Bioinformatics

https://bioinformatics.sph.harvard.edu/training

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

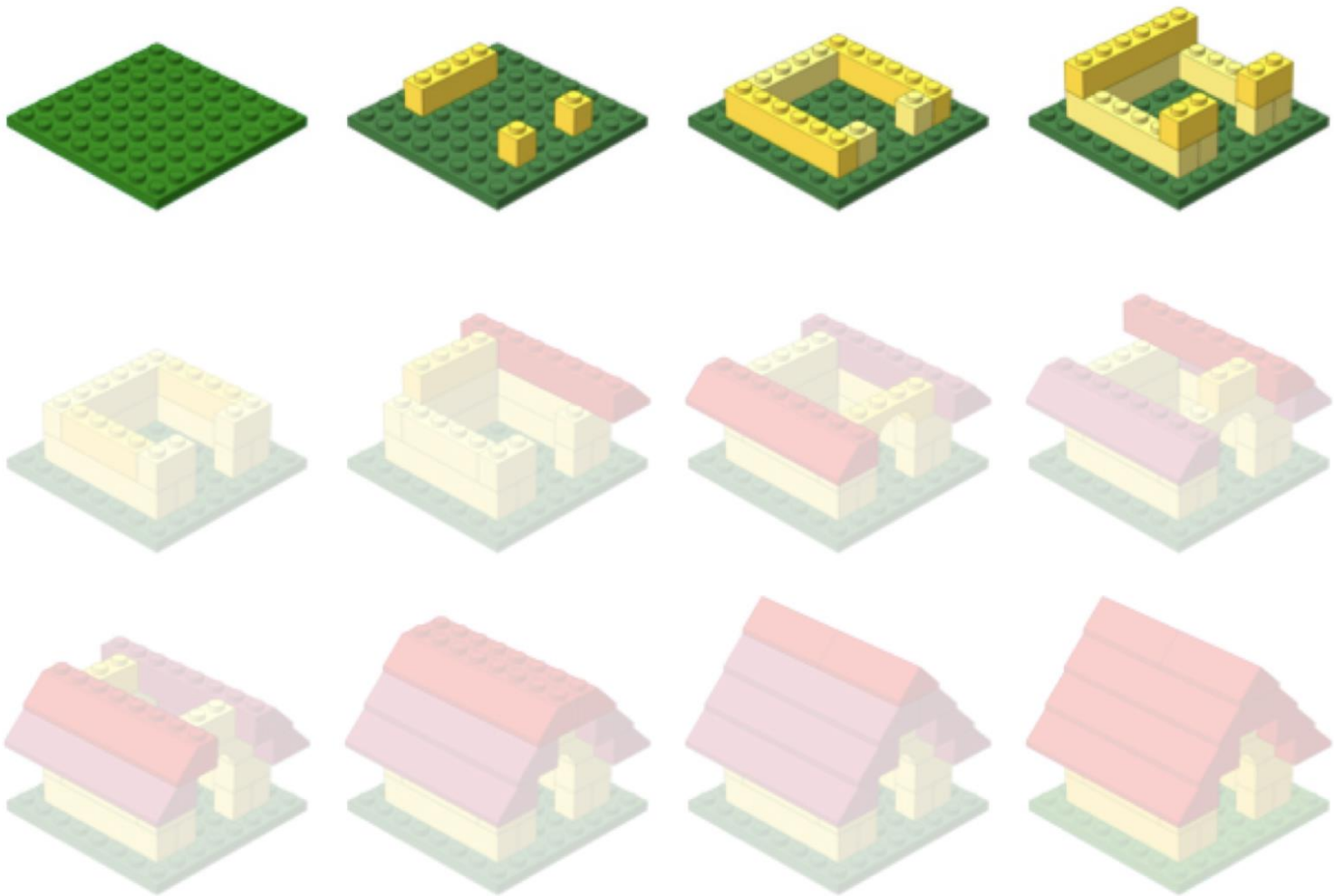
HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

Workshop scope

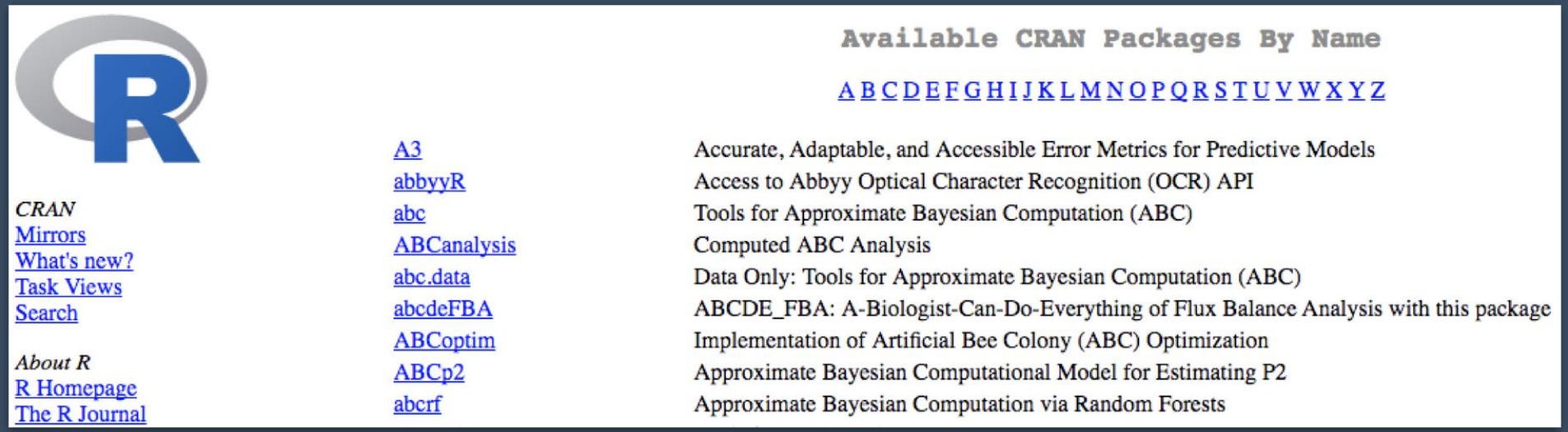


Learning R

Workshop Scope

- Comfortably use RStudio (a graphical interface for R)
- Fluently interact with R using RStudio
- Become familiar with R syntax
- Understand data structures in R
- Inspect and manipulate data structures
- Install packages and use functions in R

CRAN



CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Available CRAN Packages By Name

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abbyyR	Access to Abbyy Optical Character Recognition (OCR) API
abc	Tools for Approximate Bayesian Computation (ABC)
ABCanalysis	Computed ABC Analysis
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCoptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
abcrf	Approximate Bayesian Computation via Random Forests

- Comprehensive R Archive Network
- The main repository for R packages
- Easy to install

https://cran.r-project.org/

- An alternative package repository; “..provides tools for the analysis and comprehension of *high-throughput genomic data*.”
- Includes (but is not limited to) tools for:
- Performing statistical analysis
- Accessing public datasets
- Open source and open development
- Free

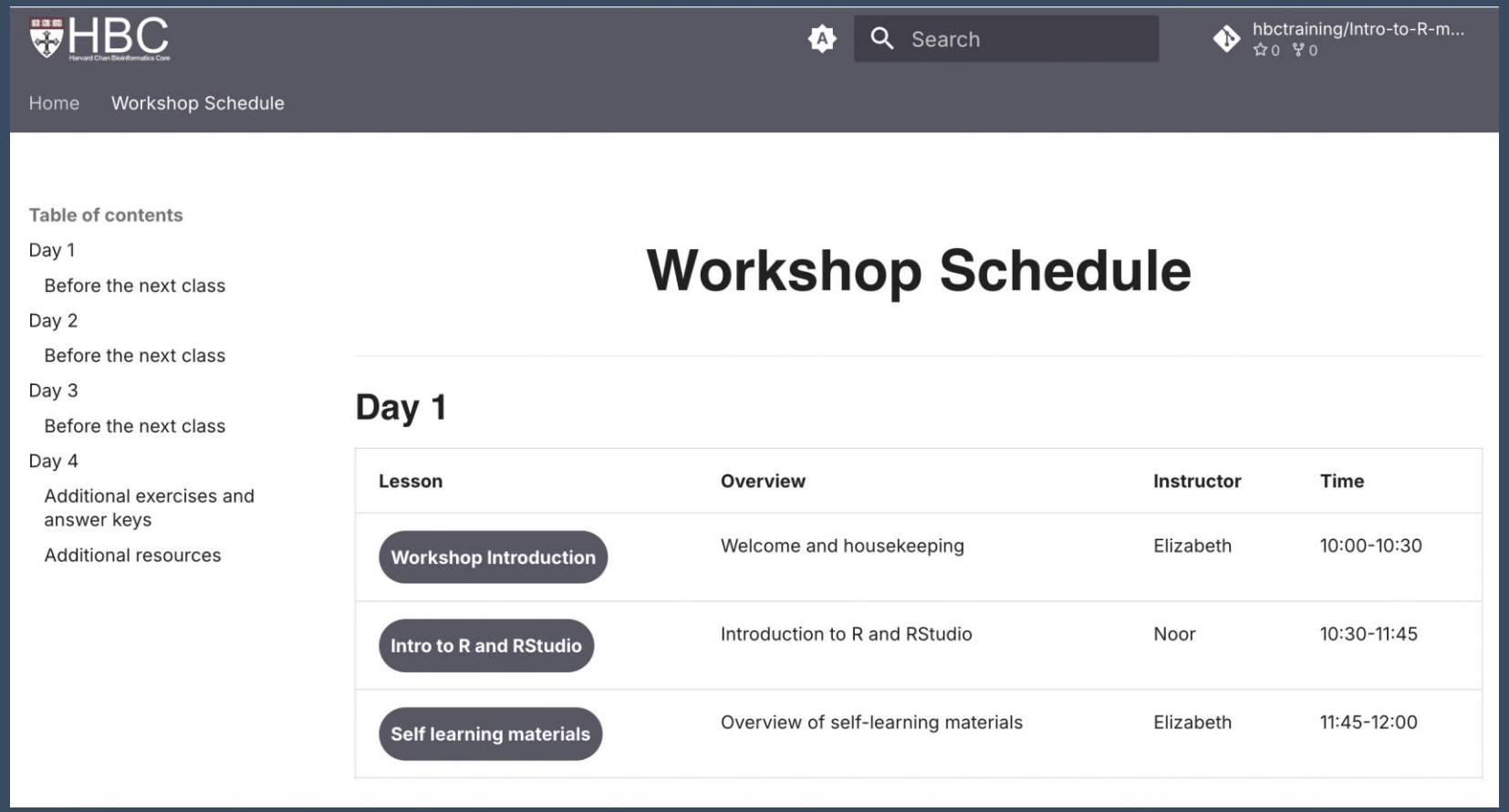
https://www.bioconductor.org

Workshop Scope

- Comfortably use RStudio (a graphical interface for R)
- Fluently interact with R using RStudio
- Become familiar with R syntax
- Understand data structures in R
- Inspect and manipulate data structures
- Install packages and use functions in R
- Visualize data using ggplot2
- Utilize pipes, tibbles and functions from the Tidyverse package suite

Logistics

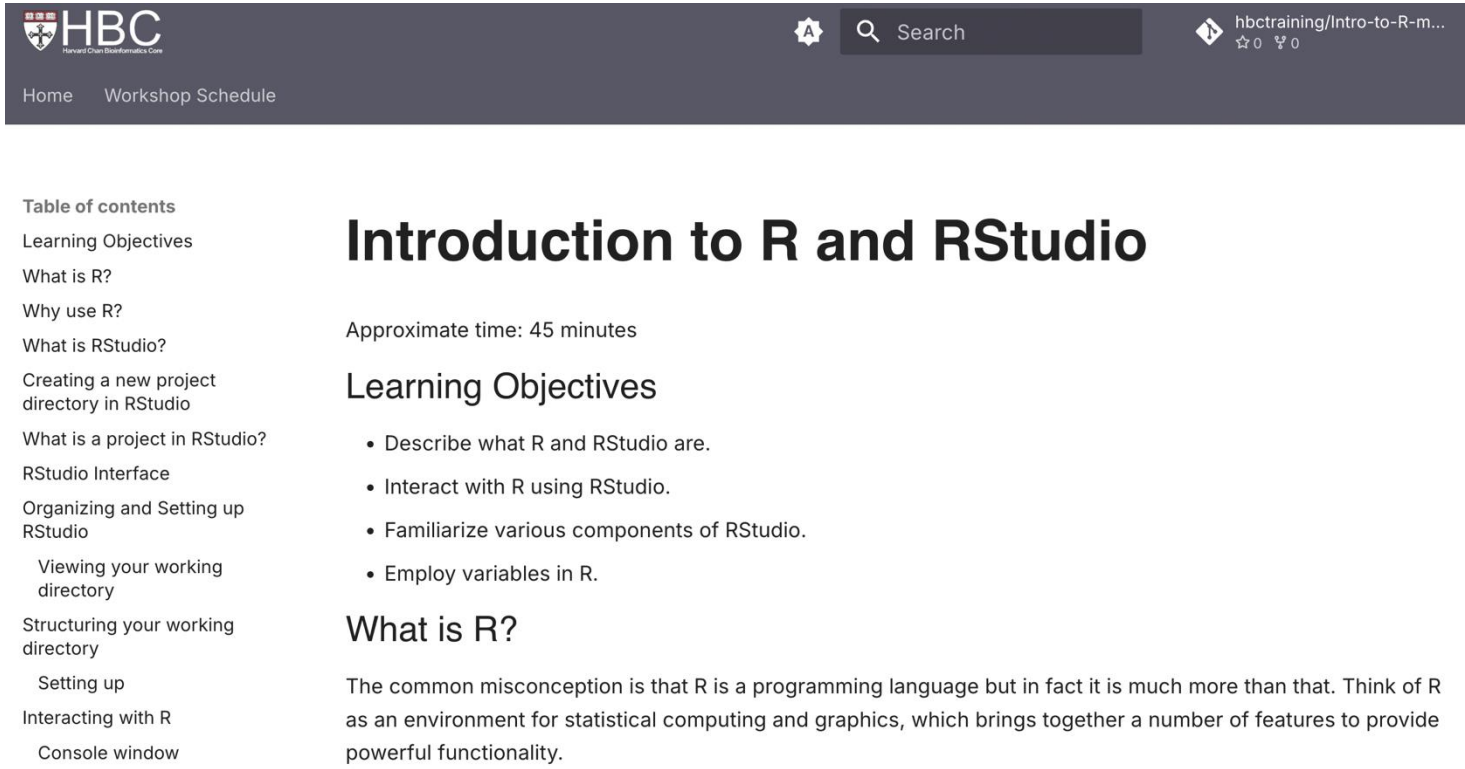
Course schedule



https://tinyurl.com/hbc-r-revised

Course materials

- We continuously update our materials to reflect changes in the field/software



https://tinyurl.com/hbc-r-revised

Course participation

- Mandatory review of self-learning lessons and assignments
- Attendance required for all classes
- Your questions and active participation drive learning
- **We look forward to all of your questions!**

Course participation

- At-home lessons and exercises after each session
- Cover material not previously discussed
- Provides us feedback to help pace the course appropriately
- 3-5 hours to complete
- Homework load is heavier in the beginning of this workshop series and tapers off

Using AI for Assignments

- Do
  - Try to resolve error messages with it
  - Test code written by AI on a dataset where you have expected results
  - Take the time to review the generated code line-by-line
- Don’t
  - Implement it in replacement to learning
  - Write code that you don’t understand
  - Assume the output from an AI process is correct

Odds & Ends

- Quit/minimize all applications that are not required for class
- Name tags
- Post-its
  - green - I am all set
  - red - I need time/help
- Phones on vibrate/silent
- Bathrooms

Contact Us

HBC
Harvard Chan Bioinformatics Core

- *HBC training team:* hbctraining@hsph.harvard.edu
- *HBC consulting:* bioinformatics@hsph.harvard.edu